

HIGH AFFINITY HUMAN ANTIBODIES AND HUMAN ANTIBODIES AGAINST
HUMAN ANTIGENS

CROSS-REFERENCE TO RELATED APPLICATION

This application is a continuation-in-part (CIP) of U.S. Serial No. (USSN) 08/544,404 filed 10 October 1995, now U.S. Patent No. 5,770,429; which is a CIP of USSN 08/352,322 filed 7 December 1994, now U.S. Patent No. 5,625,126; which is a CIP of USSN 08/209,741 filed March 9, 1994 (now abandoned); which is a CIP of USSN 08/165,699 filed December 10, 1993 (now abandoned); which is a CIP of USSN 08/161,739 filed December 3, 1993 (now abandoned); which is a CIP of USSN 08/155,301 filed November 18, 1993 (now abandoned); which is a CIP of USSN 08/096,762 filed July 22, 1993, now U.S. Patent No. 5,814,318; which is a CIP of USSN 08/053,131 filed April 26, 1993, now U.S. Patent No. 5,661,016; which is a CIP of USSN 07/990,860 filed December 16, 1992, now U.S. Patent No. 5,545,806; which is a CIP of USSN 07/904,068 filed June 23, 1992 (now abandoned), which is a CIP of USSN 07/853,408 filed March 18, 1992, now U.S. Patent No. 5,789,650; which is a CIP of USSN 07/834,539 filed February 5, 1992, now U.S. Patent No. 5,633,425; which is a CIP of USSN 07/810,279 filed December 17, 1991, now U.S. Patent No. 5,569,825; which is a CIP of USSN 07/575,962 filed August 31, 1990 (now abandoned), which is a CIP of USSN 07/574,748 filed August 29, 1990 (now abandoned).

TECHNICAL FIELD

The invention relates to transgenic non-human animals capable of producing heterologous antibodies, transgenes used to produce such transgenic animals, transgenes capable of functionally rearranging a heterologous D gene in V-D-J recombination, immortalized B-cells capable of producing heterologous antibodies, methods and transgenes for producing heterologous antibodies of multiple isotypes, methods and transgenes for producing heterologous antibodies wherein a variable region sequence comprises somatic mutation as compared to germline rearranged variable region sequences,

transgenic nonhuman animals which produce antibodies having a human primary sequence and which bind to human antigens, hybridomas made from B cells of such transgenic animals, and monoclonal antibodies expressed by such hybridomas.

BACKGROUND OF THE INVENTION

One of the major impediments facing the development of in vivo therapeutic and diagnostic applications for monoclonal antibodies in humans is the intrinsic immunogenicity of non-human immunoglobulins. For example, when immunocompetent human patients are administered therapeutic doses of rodent monoclonal antibodies, the patients produce antibodies against the rodent immunoglobulin sequences; these human anti-mouse antibodies (HAMA) neutralize the therapeutic antibodies and can cause acute toxicity. Hence, it is desirable to produce human immunoglobulins that are reactive with specific human antigens that are promising therapeutic and/or diagnostic targets. However, producing human immunoglobulins that bind specifically with human antigens is problematic.

The present technology for generating monoclonal antibodies involves pre-exposing, or priming, an animal (usually a rat or mouse) with antigen, harvesting B-cells from that animal, and generating a library of hybridoma clones. By screening a hybridoma population for antigen binding specificity (idiotype) and also screening for immunoglobulin class (isotype), it is possible to select hybridoma clones that secrete the desired antibody.

However, when present methods for generating monoclonal antibodies are applied for the purpose of generating human antibodies that have binding specificities for human antigens, obtaining B-lymphocytes which produce human immunoglobulins a serious obstacle, since humans will typically not make immune responses against self-antigens.

Hence, present methods of generating human monoclonal antibodies that are specifically reactive with human antigens are clearly insufficient. It is evident that the same limitations on generating monoclonal antibodies to

authentic self antigens apply where non-human species are used as the source of B-cells for making the hybridoma.

The construction of transgenic animals harboring a functional heterologous immunoglobulin transgene are a method by which antibodies reactive with self antigens may be produced. However, in order to obtain expression of therapeutically useful antibodies, or hybridoma clones producing such antibodies, the transgenic animal must produce transgenic B cells that are capable of maturing through the B lymphocyte development pathway. Such maturation requires the presence of surface IgM on the transgenic B cells, however isotypes other than IgM are desired for therapeutic uses. Thus, there is a need for transgenes and animals harboring such transgenes that are able to undergo functional V-D-J rearrangement to generate recombinational diversity and junctional diversity. Further, such transgenes and transgenic animals preferably include cis-acting sequences that facilitate isotype switching from a first isotype that is required for B cell maturation to a subsequent isotype that has superior therapeutic utility.

A number of experiments have reported the use of transfected cell lines to determine the specific DNA sequences required for Ig gene rearrangement (reviewed by Lewis and Gellert (1989), Cell, 59, 585-588). Such reports have identified putative sequences and concluded that the accessibility of these sequences to the recombinase enzymes used for rearrangement is modulated by transcription (Yancopoulos and Alt (1985), Cell, 40, 271-281). The sequences for V(D)J joining are reportedly a highly conserved, near-palindromic heptamer and a less well conserved AT-rich nanomer separated by a spacer of either 12 or 23 bp (Tonegawa (1983), Nature, 302, 575-581; Hesse, et al. (1989), Genes in Dev., 3, 1053-1061). Efficient recombination reportedly occurs only between sites containing recombination signal sequences with different length spacer regions.

Ig gene rearrangement, though studied in tissue culture cells, has not been extensively examined in transgenic mice. Only a handful of reports have been published

describing rearrangement test constructs introduced into mice [Buchini, et al. (1987), Nature, 326, 409-411 (unrearranged chicken λ transgene); Goodhart, et al. (1987), Proc. Natl. Acad. Sci. USA, 84, 4229-4233) (unrearranged rabbit κ gene); and Bruggemann, et al. (1989), Proc. Natl. Acad. Sci. USA, 86, 6709-6713 (hybrid mouse-human heavy chain)]. The results of such experiments, however, have been variable, in some cases, producing incomplete or minimal rearrangement of the transgene.

Further, a variety of biological functions of antibody molecules are exerted by the Fc portion of molecules, such as the interaction with mast cells or basophils through Fc ϵ , and binding of complement by Fc μ or Fc γ , it further is desirable to generate a functional diversity of antibodies of a given specificity by variation of isotype.

Although transgenic animals have been generated that incorporate transgenes encoding one or more chains of a heterologous antibody, there have been no reports of heterologous transgenes that undergo successful isotype switching. Transgenic animals that cannot switch isotypes are limited to producing heterologous antibodies of a single isotype, and more specifically are limited to producing an isotype that is essential for B cell maturation, such as IgM and possibly IgD, which may be of limited therapeutic utility. Thus, there is a need for heterologous immunoglobulin transgenes and transgenic animals that are capable of switching from an isotype needed for B cell development to an isotype that has a desired characteristic for therapeutic use.

Based on the foregoing, it is clear that a need exists for methods of efficiently producing heterologous antibodies, e.g. antibodies encoded by genetic sequences of a first species that are produced in a second species. More particularly, there is a need in the art for heterologous immunoglobulin transgenes and transgenic animals that are capable of undergoing functional V-D-J gene rearrangement that incorporates all or a portion of a D gene segment which contributes to recombinational diversity. Further, there is a need in the art for transgenes and transgenic animals that can

support V-D-J recombination and isotype switching so that (1) functional B cell development may occur, and (2) therapeutically useful heterologous antibodies may be produced. There is also a need for a source of B cells which can be used to make hybridomas that produce monoclonal antibodies for therapeutic or diagnostic use in the particular species for which they are designed. A heterologous immunoglobulin transgene capable of functional V-D-J recombination and/or capable of isotype switching could fulfill these needs.

In accordance with the foregoing object transgenic nonhuman animals are provided which are capable of producing a heterologous antibody, such as a human antibody.

Further, it is an object to provide B-cells from such transgenic animals which are capable of expressing heterologous antibodies wherein such B-cells are immortalized to provide a source of a monoclonal antibody specific for a particular antigen.

In accordance with this foregoing object, it is a further object of the invention to provide hybridoma cells that are capable of producing such heterologous monoclonal antibodies.

Still further, it is an object herein to provide heterologous unrearranged and rearranged immunoglobulin heavy and light chain transgenes useful for producing the aforementioned non-human transgenic animals.

Still further, it is an object herein to provide methods to disrupt endogenous immunoglobulin loci in the transgenic animals.

Still further, it is an object herein to provide methods to induce heterologous antibody production in the aforementioned transgenic non-human animal.

A further object of the invention is to provide methods to generate an immunoglobulin variable region gene segment repertoire that is used to construct one or more transgenes of the invention.

The references discussed herein are provided solely for their disclosure prior to the filing date of the present

application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention.

SUMMARY OF THE INVENTION

Transgenic nonhuman animals are provided which are capable of producing a heterologous antibody, such as a human antibody. Such heterologous antibodies may be of various isotypes, including: IgG1, IgG2, IgG3, IgG4, IgM, IgA1, IgA2, IgA_{sec}, IgD, or IgE. In order for such transgenic nonhuman animals to make an immune response, it is necessary for the transgenic B cells and pre-B cells to produce surface-bound immunoglobulin, particularly of the IgM (or possibly IgD) isotype, in order to effectuate B cell development and antigen-stimulated maturation. Such expression of an IgM (or IgD) surface-bound immunoglobulin is only required during the antigen-stimulated maturation phase of B cell development, and mature B cells may produce other isotypes, although only a single switched isotype may be produced at a time.

Typically, a cell of the B-cell lineage will produce only a single isotype at a time, although cis or trans alternative RNA splicing, such as occurs naturally with the μ_s (secreted μ) and μ_m (membrane-bound μ) forms, and the μ and δ immunoglobulin chains, may lead to the contemporaneous expression of multiple isotypes by a single cell. Therefore, in order to produce heterologous antibodies of multiple isotypes, specifically the therapeutically useful IgG, IgA, and IgE isotypes, it is necessary that isotype switching occur. Such isotype switching may be classical class-switching or may result from one or more non-classical isotype switching mechanisms.

The invention provides heterologous immunoglobulin transgenes and transgenic nonhuman animals harboring such transgenes, wherein the transgenic animal is capable of producing heterologous antibodies of multiple isotypes by undergoing isotype switching. Classical isotype switching occurs by recombination events which involve at least one switch sequence region in the transgene. Non-classical

isotype switching may occur by, for example, homologous recombination between human σ_{μ} and human Σ_{μ} sequences (δ -associated deletion). Alternative non-classical switching mechanisms, such as intertransgene and/or interchromosomal recombination, among others, may occur and effectuate isotype switching. Such transgenes and transgenic nonhuman animals produce a first immunoglobulin isotype that is necessary for antigen-stimulated B cell maturation and can switch to encode and produce one or more subsequent heterologous isotypes that have therapeutic and/or diagnostic utility. Transgenic nonhuman animals of the invention are thus able to produce, in one embodiment, IgG, IgA, and/or IgE antibodies that are encoded by human immunoglobulin genetic sequences and which also bind specific human antigens with high affinity.

The invention also encompasses B-cells from such transgenic animals that are capable of expressing heterologous antibodies of various isotypes, wherein such B-cells are immortalized to provide a source of a monoclonal antibody specific for a particular antigen. Hybridoma cells that are derived from such B-cells can serve as one source of such heterologous monoclonal antibodies.

The invention provides heterologous unrearranged and rearranged immunoglobulin heavy and light chain transgenes capable of undergoing isotype switching in vivo in the aforementioned non-human transgenic animals or in explanted lymphocytes of the B-cell lineage from such transgenic animals. Such isotype switching may occur spontaneously or be induced by treatment of the transgenic animal or explanted B-lineage lymphocytes with agents that promote isotype switching, such as T-cell-derived lymphokines (e.g., IL-4 and IFN $_{\gamma}$).

Still further, the invention includes methods to induce heterologous antibody production in the aforementioned transgenic non-human animal, wherein such antibodies may be of various isotypes. These methods include producing an antigen-stimulated immune response in a transgenic nonhuman animal for the generation of heterologous antibodies, particularly

heterologous antibodies of a switched isotype (i.e., IgG, IgA, and IgE).

This invention provides methods whereby the transgene contains sequences that effectuate isotype switching, so that the heterologous immunoglobulins produced in the transgenic animal and monoclonal antibody clones derived from the B-cells of said animal may be of various isotypes.

This invention further provides methods that facilitate isotype switching of the transgene, so that switching between particular isotypes may occur at much higher or lower frequencies or in different temporal orders than typically occurs in germline immunoglobulin loci. Switch regions may be grafted from various C_H genes and ligated to other C_H genes in a transgene construct; such grafted switch sequences will typically function independently of the associated C_H gene so that switching in the transgene construct will typically be a function of the origin of the associated switch regions. Alternatively, or in combination with switch sequences, δ -associated deletion sequences may be linked to various C_H genes to effect non-classical switching by deletion of sequences between two δ -associated deletion sequences. Thus, a transgene may be constructed so that a particular C_H gene is linked to a different switch sequence and thereby is switched to more frequently than occurs when the naturally associated switch region is used.

This invention also provides methods to determine whether isotype switching of transgene sequences has occurred in a transgenic animal containing an immunoglobulin transgene.

The invention provides immunoglobulin transgene constructs and methods for producing immunoglobulin transgene constructs, some of which contain a subset of germline immunoglobulin loci sequences (which may include deletions). The invention includes a specific method for facilitated cloning and construction of immunoglobulin transgenes, involving a vector that employs unique XhoI and SalI restriction sites flanked by two unique NotI sites. This method exploits the complementary termini of XhoI and SalI

restrictions sites and is useful for creating large constructs by ordered concatemerization of restriction fragments in a vector.

The transgenes of the invention include a heavy chain transgene comprising DNA encoding at least one variable gene segment, one diversity gene segment, one joining gene segment and one constant region gene segment. The immunoglobulin light chain transgene comprises DNA encoding at least one variable gene segment, one joining gene segment and one constant region gene segment. The gene segments encoding the light and heavy chain gene segments are heterologous to the transgenic non-human animal in that they are derived from, or correspond to, DNA encoding immunoglobulin heavy and light chain gene segments from a species not consisting of the transgenic non-human animal. In one aspect of the invention, the transgene is constructed such that the individual gene segments are unrearranged, i.e., not rearranged so as to encode a functional immunoglobulin light or heavy chain. Such unrearranged transgenes permit recombination of the gene segments (functional rearrangement) and expression of the resultant rearranged immunoglobulin heavy and/or light chains within the transgenic non-human animal when said animal is exposed to antigen.

In one aspect of the invention, heterologous heavy and light immunoglobulin transgenes comprise relatively large fragments of unrearranged heterologous DNA. Such fragments typically comprise a substantial portion of the C, J (and in the case of heavy chain, D) segments from a heterologous immunoglobulin locus. In addition, such fragments also comprise a substantial portion of the variable gene segments.

In one embodiment, such transgene constructs comprise regulatory sequences, e.g. promoters, enhancers, class switch regions, recombination signals and the like, corresponding to sequences derived from the heterologous DNA. Alternatively, such regulatory sequences may be incorporated into the transgene from the same or a related species of the non-human animal used in the invention. For example, human immunoglobulin gene segments may be combined in a transgene

with a rodent immunoglobulin enhancer sequence for use in a transgenic mouse.

In a method of the invention, a transgenic non-human animal containing germline unrearranged light and heavy immunoglobulin transgenes - that undergo VDJ joining during D-cell differentiation - is contacted with an antigen to induce production of a heterologous antibody in a secondary repertoire B-cell.

Also included in the invention are vectors and methods to disrupt the endogenous immunoglobulin loci in the non-human animal to be used in the invention. Such vectors and methods utilize a transgene, preferably positive-negative selection vector, which is constructed such that it targets the functional disruption of a class of gene segments encoding a heavy and/or light immunoglobulin chain endogenous to the non-human animal used in the invention. Such endogenous gene segments include diversity, joining and constant region gene segments. In this aspect of the invention, the positive-negative selection vector is contacted with at least one embryonic stem cell of a non-human animal after which cells are selected wherein the positive-negative selection vector has integrated into the genome of the non-human animal by way of homologous recombination. After transplantation, the resultant transgenic non-human animal is substantially incapable of mounting an immunoglobulin-mediated immune response as a result of homologous integration of the vector into chromosomal DNA. Such immune deficient non-human animals may thereafter be used for study of immune deficiencies or used as the recipient of heterologous immunoglobulin heavy and light chain transgenes.

The invention also provides vectors, methods, and compositions useful for suppressing the expression of one or more species of immunoglobulin chain(s), without disrupting an endogenous immunoglobulin locus. Such methods are useful for suppressing expression of one or more endogenous immunoglobulin chains while permitting the expression of one or more transgene-encoded immunoglobulin chains. Unlike genetic disruption of an endogenous immunoglobulin chain

locus, suppression of immunoglobulin chain expression does not require the time-consuming breeding that is needed to establish transgenic animals homozygous for a disrupted endogenous Ig locus. An additional advantage of suppression as compared to endogenous Ig gene disruption is that, in certain embodiments, chain suppression is reversible within an individual animal. For example, Ig chain suppression may be accomplished with: (1) transgenes encoding and expressing antisense RNA that specifically hybridizes to an endogenous Ig chain gene sequence, (2) antisense oligonucleotides that specifically hybridize to an endogenous Ig chain gene sequence, and (3) immunoglobulins that bind specifically to an endogenous Ig chain polypeptide.

The invention provides transgenic non-human animals comprising: a homozygous pair of functionally disrupted endogenous heavy chain alleles, a homozygous pair of functionally disrupted endogenous light chain alleles, at least one copy of a heterologous immunoglobulin heavy chain transgene, and at least one copy of a heterologous immunoglobulin heavy chain transgene, wherein said animal makes an antibody response following immunization with an antigen, such as a human antigen (e.g., CD4). The invention also provides such a transgenic non-human animal wherein said functionally disrupted endogenous heavy chain allele is a J_H region homologous recombination knockout, said functionally disrupted endogenous light chain allele is a J_K region homologous recombination knockout, said heterologous immunoglobulin heavy chain transgene is the HC1 or HC2 human minigene transgene, said heterologous light chain transgene is the KC2 or KC1e human κ transgene, and wherein said antigen is a human antigen.

The invention also provides various embodiments for suppressing, ablating, and/or functionally disrupting the endogenous nonhuman immunoglobulin loci.

The invention also provides transgenic mice expressing both human sequence heavy chains and chimeric heavy chains comprising a human sequence heavy chain variable region and a murine sequence heavy chain constant region. Such

chimeric heavy chains are generally produced by trans-switching between a functionally rearranged human transgene and an endogenous murine heavy chain constant region (e.g., $\gamma 1$, $\gamma 2a$, $\gamma 2b$, $\gamma 3$). Antibodies comprising such chimeric heavy chains, typically in combination with a transgene-encoded human sequence light chain or endogenous murine light chain, are formed in response to immunization with a predetermined antigen. The transgenic mice of these embodiments can comprise B cells which produce (express) a human sequence heavy chain at a first timepoint and trans-switch to produce (express) a chimeric heavy chain composed of a human variable region and a murine constant region (e.g., $\gamma 1$, $\gamma 2a$, $\gamma 2b$, $\gamma 3$) at a second (subsequent) timepoint; such human sequence and chimeric heavy chains are incorporated into functional antibodies with light chains; such antibodies are present in the serum of such transgenic mice. Thus, to restate: the transgenic mice of these embodiments can comprise B cells which express a human sequence heavy chain and subsequently switch (via trans-switching or cis-switching) to express a chimeric or isotype-switched heavy chain composed of a human variable region and an alternative constant region (e.g., murine $\gamma 1$, $\gamma 2a$, $\gamma 2b$, $\gamma 3$; human γ , α , ϵ); such human sequence and chimeric or isotype-switched heavy chains are incorporated into functional antibodies with light chains (human or mouse); such antibodies are present in the serum of such transgenic mice.

The invention also provides a method for generating a large transgene, said method comprising:

introducing into a mammalian cell at least three polynucleotide species; a first polynucleotide species having a recombinogenic region of sequence identity shared with a second polynucleotide species, a second polynucleotide species having a recombinogenic region of sequence identity shared with a first polynucleotide species and a recombinogenic region of sequence identity shared with a third polynucleotide species, and a third polynucleotide species having a recombinogenic region of sequence identity shared with said second polynucleotide species.

Recombinogenic regions are regions of substantial sequence identity sufficient to generate homologous recombination in vivo in a mammalian cell (e.g., ES cell), and preferably also in non-mammalian eukaryotic cells (e.g., *Saccharomyces* and other yeast or fungal cells). Typically, recombinogenic regions are at least 50 to 100000 nucleotides long or longer, preferably 500 nucleotides to 10000 nucleotides long, and are often 80-100 percent identical, frequently 95-100 percent identical, often isogenic.

The references discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 depicts the complementarity determining regions CDR1, CDR2 and CDR3 and framework regions FR1, FR2, FR3 and FR4 in unrearranged genomic DNA and mRNA expressed from a rearranged immunoglobulin heavy chain gene,

Fig. 2 depicts the human λ chain locus,

Fig. 3 depicts the human κ chain locus,

Fig. 4 depicts the human heavy chain locus,

Fig. 5 depicts a transgene construct containing a rearranged IgM gene ligated to a 25 kb fragment that contains human $\gamma 3$ and $\gamma 1$ constant regions followed by a 700 bp fragment containing the rat chain 3' enhancer sequence.

Fig. 6 is a restriction map of the human κ chain locus depicting the fragments to be used to form a light chain transgene by way of in vivo homologous recombination.

Fig. 7 depicts the construction of pGP1.

Fig. 8 depicts the construction of the polylinker contained in pGP1.

Fig. 9 depicts the fragments used to construct a human heavy chain transgene of the invention.

Fig. 10 depicts the construction of pHIG1 and pCON1.

Fig. 11 depicts the human C γ 1 fragments which are inserted into pRE3 (rat enhancer 3') to form pREG2.

Fig. 12 depicts the construction of pHIG3' and PCON.

Fig. 13 depicts the fragment containing human D region segments used in construction of the transgenes of the invention.

Fig. 14 depicts the construction of pHIG2 (D segment containing plasmid).

Fig. 15 depicts the fragments covering the human Jk and human Ck gene segments used in constructing a transgene of the invention.

Fig. 16 depicts the structure of pEp.

Fig. 17 depicts the construction of pKaph.

Figs. 18A through 18D depict the construction of a positive-negative selection vector for functionally disrupting the endogenous heavy chain immunoglobulin locus of mouse.

Figs. 19A through 19C depict the construction of a positive-negative selection vector for functionally disrupting the endogenous immunoglobulin light chain loci in mouse.

Figs. 20A through 20E depict the structure of a kappa light chain targeting vector.

Figs. 21A through 21F depict the structure of a mouse heavy chain targeting vector.

Fig. 22 depicts the map of vector pGPe.

Fig. 23 depicts the structure of vector pJM2.

Fig. 24 depicts the structure of vector pCOR1.

Fig. 25 depicts the transgene constructs for pIGM1, pHCl and pHCl2.

Fig. 26 depicts the structure of pye2.

Fig. 27 depicts the structure of pVGE1.

Fig. 28 depicts the assay results of human Ig expression in a pHCl transgenic mouse.

Fig. 29 depicts the structure of pJCK1.

Fig. 30 depicts the construction of a synthetic heavy chain variable region.

Fig. 31 is a schematic representation of the heavy chain minilocus constructs pIGM1, pHCl, and pHCl2.

Fig. 32 is a schematic representation of the heavy chain minilocus construct pIGG1 and the κ light chain minilocus construct pKCl, pKVe1, and pKCl2.

Fig. 33 depicts a scheme to reconstruct functionally rearranged light chain genes.

Fig. 34 depicts serum ELISA results

Fig. 35 depicts the results of an ELISA assay of serum from 8 transgenic mice.

Fig. 36 is a schematic representation of plasmid pBCE1.

Figs. 37A-37C depict the immune response of transgenic mice of the present invention against KLH-DNP, by measuring IgG and IgM levels specific for KLH-DNP (37A), KLH (37B) and BSA-DNP (37C).

Fig. 38 shows ELISA data demonstrating the presence of antibodies that bind human carcinoembryonic antigen (CEA) and comprise human μ chains; each panel shows reciprocal serial dilutions from pooled serum samples obtained from mice on the indicated day following immunization.

Fig. 39 shows ELISA data demonstrating the presence of antibodies that bind human carcinoembryonic antigen (CEA) and comprise human γ chains; each panel shows reciprocal serial dilutions from pooled serum samples obtained from mice on the indicated day following immunization.

Fig. 40 shows aligned variable region sequences of 23 randomly-chosen cDNAs (SEQ ID NOS:271, 273, 274, 276-289, 291, 292 and 294-297) generated from mRNA obtained from lymphoid tissue of HCl transgenic mice immunized with human carcinoembryonic antigen (CEA) as compared to the germline transgene sequence (top line) (SEQ ID NO:269); on each line nucleotide changes relative to germline sequence are shown. The regions corresponding to heavy chain CDR1, CDR2, and CDR3 are indicated. Non-germline encoded nucleotides are shown in capital letters. J segments= SEQ ID NOS:270, 272, 275, 290 and 293.

Fig. 41 shows the nucleotide sequence of a human DNA fragment, designated vk65.3, containing a V_k gene segment (SEQ ID NO:298); the deduced amino acid sequences of the V_k coding regions are also shown (SEQ ID NO:299); splicing and recombination signal sequences (heptamer/nonamer) are shown boxed.

Fig. 42 shows the nucleotide sequence of a human DNA fragment, designated vk65.5, containing a V_k gene segment (SEQ ID NO:300); the deduced amino acid sequences of the V_k coding regions are also shown (SEQ ID NO:301); splicing and recombination signal sequences (heptamer/nonamer) are shown boxed.

Fig. 43 shows the nucleotide sequence of a human DNA fragment, designated vk65.8, containing a V_k gene segment (SEQ ID NO:302); the deduced amino acid sequences of the V_k coding regions are also shown (SEQ ID NO:303); splicing and recombination signal sequences (heptamer/nonamer) are shown boxed.

Fig. 44 shows the nucleotide sequence of a human DNA fragment, designated vk65.15, containing a V_k gene segment (SEQ ID NO:304); the deduced amino acid sequences of the V_k coding regions are also shown (SEQ ID NO:305); splicing and recombination signal sequences (heptamer/nonamer) are shown boxed.

Fig. 45 shows formation of a light chain minilocus by homologous recombination between two overlapping fragments which were co-injected.

Fig. 46 shows ELISA results for monoclonal antibodies reactive with CEA and non-CEA antigens showing the specificity of antigen binding.

Fig. 47 shows the DNA sequences of 10 cDNAs (SEQ ID NOS:306-315) amplified by PCR to amplify transcripts having a human VDJ and a murine constant region sequence.

Fig. 48 shows ELISA results for various dilutions of serum obtained from mice bearing both a human heavy chain minilocus transgene and a human κ minilocus transgene; the mouse was immunized with human CD4 and the data shown represents antibodies reactive with human CD4 and possessing human κ , human μ , or human γ epitopes, respectively.

Fig. 49 shows relative distribution of lymphocytes staining for human μ or mouse μ as determined by FACS for three mouse genotypes.

Fig. 50 shows relative distribution of lymphocytes staining for human κ or mouse κ as determined by FACS for three mouse genotypes.

Fig. 51 shows relative distribution of lymphocytes staining for mouse λ as determined by FACS for three mouse genotypes.

Fig. 52 shows relative distribution of lymphocytes staining for mouse λ or human κ as determined by FACS for four mouse genotypes.

Fig. 53 shows the amounts of human μ , human γ , human κ , mouse μ , mouse γ , mouse κ , and mouse λ chains in the serum of unimmunized 0011 mice.

Fig. 54 shows a scatter plot showing the amounts of human μ , human γ , human κ , mouse μ , mouse γ , mouse κ , and mouse λ chains in the serum of unimmunized 0011 mice of various genotypes.

Fig. 55 shows the titres of antibodies comprising human μ , human γ , or human κ chains in anti-CD4 antibodies in the serum taken at three weeks or seven weeks post-immunization following immunization of a 0011 mouse with human CD4.

Fig. 56 shows a schematic representation of the human heavy chain minilocus transgenes pHCl and pHCl2, and the light chain minilocus transgenes pKCl, pKClc, and the light chain minilocus transgene created by homologous recombination between pKCl2 and Co4 at the site indicated.

Fig. 57 shows a linkage map of the murine lambda light chain locus as taken from Storb et al. (1989) op.cit.; the stippled boxes represent a pseudogene.

Fig. 58 shows a schematic representation of inactivation of the murine λ locus by homologous gene targeting.

Fig. 59 schematically shows the structure of a homologous recombination targeting transgene for deleting genes, such as heavy chain constant region genes.

Fig. 60 shows a map of the BALB/c murine heavy chain locus as taken from Immunoglobulin Genes, Honjo, T, Alt, FW, and Rabbits TH (eds.) Academic Press, NY (1989) p. 129.

Structural genes are shown by closed boxes in the top line; second and third lines show restriction sites with symbols indicated.

Fig. 61 shows a nucleotide sequence (SEQ ID NO:316) of mouse heavy chain locus α constant region gene.

Fig. 62 shows the construction of a frameshift vector (plasmid B) for introducing a two bp frameshift into the murine heavy chain locus J_4 gene.

Fig. 63 shows isotype specific response of transgenic animals during hyperimmunization. The relative levels of reactive human μ and $\gamma 1$ are indicated by a colorimetric ELISA assay (y-axis). We immunized three 7-10 week old male HC1 line 57 transgenic animals (#1991, #2356, #2357), in a homozygous JHD background, by intraperitoneal injections of CEA in Freund's adjuvant. The figure depicts binding of 250 fold dilutions of pooled serum (collected prior to each injection) to CEA coated microtiter wells.

Fig. 64A and 64B show expression of transgene encoded $\gamma 1$ isotype mediated by class switch recombination. The genomic structure of integrated transgenes in two different human $\gamma 1$ expressing hybridomas is consistent with recombination between the μ and $\gamma 1$ switch regions. Fig. 64A shows a Southern blot of $PacI/SfiI$ digested DNA isolated from three transgene expressing hybridomas. From left to right: clone 92-09A-5H1-5, human $\gamma 1^+/\mu^-$; clone 92-90A-4G2-2, human $\gamma 1^+/\mu^-$; clone 92-09A-4F7-A5-2, human $\gamma 1^-, \mu^+$. All three hybridomas are derived from a 7 month old male mouse hemizygous for the HC1-57 integration, and homozygous for the JHD disruption (mouse #1991). The blot is hybridized with a probe derived from a 2.3 kb $BglII/SfiI$ DNA fragment spanning the 3' half of the human $\gamma 1$ switch region. No switch product is found in the μ expressing hybridoma, while the two $\gamma 1$ expressing hybridomas, 92-09A-5H1-5 and 92-09A-4G2-2, contain switch products resulting in $PacI/SfiI$ fragments of 5.1 and 5.3 kb respectively. Fig. 64B is a diagram of two possible deletional mechanisms by which a class switch from μ to $\gamma 1$ can occur. The human μ gene is flanked by 400 bp direct repeats ($\sigma\mu$ and $\Sigma\mu$) which can recombine to delete μ . Class switching

by this mechanism will always generate a 6.4 kb PacI/SfiI fragment, while class switching by recombination between the μ and the $\gamma 1$ switch regions will generate a PacI/SfiI fragment between 4 and 7 kb, with size variation between individual switch events. The two $\gamma 1$ expressing hybridomas examined in Fig. 64A appear to have undergone recombination between the μ and $\gamma 1$ switch regions.

Fig.65 shows chimeric human/mouse immunoglobulin heavy chains generated by trans-switching. cDNA clones of trans-switch products were generated by reverse transcription and PCR amplification of a mixture of spleen and lymph node RNA isolated from a hyperimmunized HC1 transgenic-JHD mouse (#2357; see legend to Fig.63 for description of animal and immunization schedule). The partial nucleotide sequence of 10 randomly picked clones is shown (SEQ ID NOS:317-326). Lower case letters indicate germline encoded, capital letters indicate nucleotides that cannot be assigned to known germline sequences; these may be somatic mutations, N nucleotides, or truncated D segments. Both face type indicates mouse γ sequences.

Figs.66A and 66B show that the rearranged VH251 transgene undergoes somatic mutation in a hyperimmunized. The partial nucleotide sequence of IgG heavy chain variable region cDNA clones from CH1 line 26 mice exhibiting Fig.66A primary and Fig.66B secondary responses to antigen. Germline sequence is shown at the top (SEQ ID NO:269); nucleotide changes from germline are given for each clone. A period indicates identity with germline sequence, capital letters indicate no identified germline origin. The sequences are grouped according to J segment usage. The germline sequence of each of the J segments is shown (SEQ ID NOS:327, 331, 333, 340, 345, 347, 333 and 335). Lower case letters within CDR3 sequences indicate identity to known D segment included in the HC1 transgene. The assigned D segments are indicated at the end of each sequence. Unassigned sequences could be derived from N region addition or somatic mutation; or in some cases they are simply too short to distinguish random N nucleotides from known D segments. Fig. 66A primary response: 13

randomly picked VH251- γ 1 cDNA clones (SEQ ID NOS:328-330, 332, 334-339, 341, and 342). A 4 week old female HC1 line 26-JHD mouse (#2599) was given a single injection of KLH and complete Freund's adjuvant; spleen cell RNA was isolated 5 days later. The overall frequency of somatic mutations within the V segment is 0.06% (2/3,198 bp). Fig.66B secondary response: 13 randomly picked VH251- γ 1 cDNA clones (SEQ ID NOS:344, 346, 348-352, 354, and 356). A 2 month old female HC1 line 26-JHD mouse (#3204) was given 3 injections of HEL and Freund's adjuvant over one month (a primary injection with complete adjuvant and boosts with incomplete at one week and 3 weeks); spleen and lymph node RNA was isolated 4 months later. The overall frequency of somatic mutations within the V segment is 1.6% (52/3,198 bp).

Figs. 67A and 67B show that extensive somatic mutation is confined to γ 1 sequences: somatic mutation and class switching occur within the same population of B cells. Partial nucleotide sequence of VH251 cDNA clones isolated from spleen and lymph node cells of HC1 line 57 transgenic-JHD mouse (#2357) hyperimmunized against CEA (see Fig.63 for immunization schedule). Fig.67A: IgM: 24 [~~23~~] randomly picked VH251- μ cDNA clones (SEQ ID NOS:357-361). Nucleotide sequence of 156 bp segment including CDRs 1 and 2 surrounding residues (germline sequence=SEQ ID NO:257.). The overall level of somatic mutation is 0.1% (5/3,744 bp). Fig 67B: IgG: 23 randomly picked VH251- γ 1 cDNA clones (SEQ ID NO:362-384). Nucleotide sequence of segment including CDRs 1 through 3 and surrounding residues (germline sequence=SEQ ID NO:269; J segments= SEQ ID NOS:343, 345, 347, 353 and 355). The overall frequency of somatic mutation within the V segment is 1.1% (65/5,658 bp). For comparison with the μ sequences in Fig. 67A: the mutation frequency for first 156 nucleotides is 1.1% (41/3,588 bp). See legend to Figs.66A and 66B for explanation of symbols.

Fig.68 indicates that VH51P1 and VH56P1 show extensive somatic mutation of in an unimmunized mouse. The partial nucleotide sequence of IgG heavy chain variable region cDNA clones (SEQ ID NOS:386-402, 404, and 406-408) from a 9

week old, unimmunized female HC2 line 2550 transgenic-JHD mouse (#5250) (germline sequences=SEQ ID NOS:385, 403 and 405; J segments= SEQ ID NOS:343, 345, 347, 353 and 355). The overall frequency of somatic mutation with the 19 VH56p1 segments is 2.2% (101/4,674 bp). The overall frequency of somatic mutation within the single VH51p1 segment is 2.0% (5/246 bp). See legend to Figs. 66A and 66B for explanation of symbols.

Fig. 69. Double transgenic mice with disrupted endogenous Ig loci contain human IgMk positive B cells. FACS of cells isolated from spleens of 4 mice with different genotypes. Left column: control mouse (#9944, 6 wk old female JH+/-, JCK+/-; heterozygous wild-type mouse heavy and k-light chain loci, non-transgenic). Second column: human heavy chain transgenic (#9877, 6 wk old female JH-/-, JCK-/-, HC2 line 2550 +; homozygous for disrupted mouse heavy and k-light chain loci, hemizygous for HC2 transgene). Third column: human k-light chain transgenic (#9878, 6 wk old female JH-/-, JCK-/-, KCo4 line 4437 +; homozygous for disrupted mouse heavy and k-light chain loci, hemizygous for KCo4 transgene). Right column: double transgenic (#9879, 6 wk old female JH-/-m JCK-/-, HC2 line 2550 +, KCo4 line 4437 +; homozygous for disrupted mouse heavy and k-light chain loci, hemizygous for HC2 and KCo4 transgenes). Top row: spleen cells stained for expression of mouse λ light chain (x-axis) and human k light chain (y-axis). Second row: spleen cells stained for expression of human μ heavy chain (x-axis) and human k light chain (y-axis). Third row: spleen cells stained for expression of mouse μ heavy chain (x-axis) and mouse k light chain (y-axis). Bottom row: histogram of spleen cells stained for expression of mouse B220 antigen (log fluorescence: x-axis; cell number: y-axis). For each of the two color panels, the relative number of cells in each of the displayed quadrants is given as percent of a e-parameter gate based on propidium iodide staining and light scatter. The fraction of B220+ cells in each of the samples displayed in the bottom row is given as a percent of the lymphocyte light scatter gate.

Fig. 70. Secreted immunoglobulin levels in the serum of double transgenic mice. Human μ , γ , and κ , and mouse γ and λ from 18 individual HC2/KCo4 double transgenic mice homozygous for endogenous heavy and κ -light chain locus disruption. Mice: (+) HC2 line 2550 (~5 copies of HC2 per integration), KCo4 line 4436 (1-2 copies of KCo4 per integration); (O) HC2 line 2550, KCo4 line 4437 (~10 copies of KCo4 per integration); (x) HC2 line 2550, KCo4 line 4583 (~5 copies of KCo4 per integration); (\square) HC2 line 2572 (30-50 copies of HC2 per integration, KCo4 line 4437; ($_$) HC2 line 5467 (20-30 copies of HC2 per integration, KCo4 line 4437.

Figs. 71A and 71B show human antibody responses to human antigens. Fig. 71A: Primary response to recombinant human soluble CD4. Levels of human IgM and human κ light chain are reported for prebleed (O) and post-immunization (\bullet) serum from four double transgenic mice. Fig. 71B: Switching to human IgG occurs *in vivo*. Human IgG (circles) was detected with peroxidase conjugated polyclonal anti-human IgG used in the presence of 1.5 μ /ml excess IgE, κ and 1% normal mouse serum to inhibit non-specific cross-reactivity. Human κ light chain (squares) was detected using a peroxidase conjugated polyclonal anti-human κ reagent in the presence of 1% normal mouse serum. A representative result from one mouse (#9344; HC2 line 2550, KCo4 line 4436) is shown. Each point represents an average of duplicate wells minus background absorbance.

Fig. 72 shows FACS analysis of human PBL with a hybridoma supernatant that discriminates human CD4⁺ lymphocytes from human CD8⁺ lymphocytes.

Fig. 73 shows human α -CD4 IgM and IgG in transgenic mouse serum.

Fig. 74 shows competition binding experiments comparing a transgenic mouse α -human CD4 hybridoma monoclonal, 2C11-8, to the RPA-TA and Leu-3A monoclonals.

Fig. 75 shows production data for Ig expression of cultured 2C11-8 hybridoma.

Fig. 76 shows an overlapping set of plasmid inserts constituting the HCo7 transgene.

Fig. 77A depicts the nucleotide sequence (SEQ ID NO:409) of pGP2b plasmid vector.

Fig. 77B depicts the restriction map of pGP2b plasmid vector.

Fig. 78 (parts A and B) depicts cloning strategy for assembling large transgenes.

Fig. 79 shows that large inserts are unstable in high-copy pUC derived plasmids.

Fig. 80 shows phage P1 clone P1-570. Insert spans portion of human heavy chain constant region covering $\gamma 3$ and $\gamma 1$, together with switch elements. N, NotI; S, SalI, X, XhoI.

Fig. 81 shows serum expression of human μ and $\gamma 1$ in HCo7 transgenic founder animals.

Fig. 82 shows serum expression of human immunoglobulins in HCo7/KCo4 double transgenic/double deletion mice.

Fig. 83 shows RT PCR detection of human $\gamma 1$ and $\gamma 3$ transcripts in HCo7 transgenic mouse spleen RNA.

Fig. 84 shows induction of human IgG1 and IgG3 by LPS and IL-4 in vitro.

Fig. 85. Agarose gel electrophoresis apparatus for concentration of YAC DNA.

Fig. 86. Two color FACS analysis of bone marrow cells from HC2/KCo5/JHD/JKD and HC2/KCo4/JHD/JKD mice. The fraction of cells in each of the $B220^+/CD43^-$, $B220^+/CD43^+$, and $B220^+/IgM^+$ gates is given as a percent.

Fig. 87. Two color FACS analysis of spleen cells from HC2/KCo5/JHD/JKD and HC2/KCo4/JHD/JKD mice. The fraction of cells in each of the $B220^{bright}/IgM^+$ and $B220^{dull}/IgM^+$ gates is given as a percent.

Fig. 88. Binding of IgGk anti-nCD4 monoclonal antibodies to CD4+ SupT1 cells.

Fig. 89 Epitope determination for IgG anti-nCD4 monoclonal antibodies by flow cytometry. SupT1 cells were pre-incubated with buffer (left column), 2.5 mg/ml RPA-T4 (middle column), or 2.5 mg/ml Leu3a (right column) and then with one of the 10 human IgG monoclonal antibodies (in supernatant diluted 1:2), or chimeric Leu3a. Results for 3

representative human IgG monoclonal antibodies are shown in this figure.

Fig. 90 Inhibition of an MLR by a human IgGk anti-CD4 monoclonal antibody.

Table 1 depicts the sequence of vector pGPe.

Table 2 depicts the sequence of gene V_H49.8.

Table 3 depicts the detection of human IgM and IgG in the serum of transgenic mice of this invention.

Table 4 depicts sequences of VDJ joints.

Table 5 depicts the distribution of J segments incorporated into pHCl transgene encoded transcripts to J segments found in adult human peripheral blood lymphocytes (PBL).

Table 6 depicts the distribution of D segments incorporated into pHCl transgene encoded transcripts to D segments found in adult human peripheral blood lymphocytes (PBL).

Table 7 depicts the predicted amino acid sequences of the VDJ regions from 30 clones analyzed from pHCl transgenic.

Table 8 depicts the length of the CDR3 peptides from transcripts with in-frame VDJ joints in the pHCl transgenic mouse and in human PBL.

Table 9 shows transgenic mice of line 112 that were used in the indicated experiments; (+) indicates the presence of the respective transgene, (++) indicates that the animal is homozygous for the J_HD knockout transgene.

Table 10 shows the genotypes of several 0011 mice.

Table 11 shows human variable region usage in hybridomas from transgenic mice.

Table 12 shows transgene V and J segment usage.

Table 13 shows the occurrence of somatic mutation in the HC2 heavy chain transgene in transgenic mice.

Table 14 shows identification of human V_k segments on the YAC 4x17E1.

Table 15. Identification of human V_k genes expressed in mouse line KCo5-9272.

Table 16. Secretion levels for human IgGk Anti-nCD4 monoclonal antibodies

Table 17. Rate and affinity constants for monoclonal antibodies that bind to human CD4.

Table 18. Affinity and rate constants of human anti-human CD4 monoclonal antibodies.

Table 19. Avidity and rate constants of human anti-human CD4 monoclonal antibodies.

Table 20. Avidity and rate constants reported for anti CD4 monoclonal antibodies.

Table 21. Avidity constants of human anti-human CD4 monoclonal antibodies as determined by flow cytometry.

Table 22. Partial Nucleotide Sequence for Functional Transcripts.

Table 23 Germline V(D)J Segment Usage in Hybridoma Transcripts.

Table 24. Primers, Vectors and Products Used in Minigene Construction.

Table 25. Effect of Human mAbs on Peripheral Chimpanzee Lymphocytes.

DETAILED DESCRIPTION

As has been discussed supra, it is desirable to produce human immunoglobulins that are reactive with specific human antigens that are promising therapeutic and/or diagnostic targets. However, producing human immunoglobulins that bind specifically with human antigens is problematic.

First, the immunized animal that serves as the source of B cells must make an immune response against the presented antigen. In order for an animal to make an immune response, the antigen presented must be foreign and the animal must not be tolerant to the antigen. Thus, for example, if it is desired to produce a human monoclonal antibody with an idiotype that binds to a human protein, self-tolerance will prevent an immunized human from making a substantial immune response to the human protein, since the only epitopes of the antigen that may be immunogenic will be those that result from polymorphism of the protein within the human population (allogeneic epitopes).

Second, if the animal that serves as the source of B-cells for forming a hybridoma (a human in the illustrative given example) does make an immune response against an authentic self antigen, a severe autoimmune disease may result in the animal. Where humans would be used as a source of B-cells for a hybridoma, such autoimmunization would be considered unethical by contemporary standards. Thus, developing hybridomas secreting human immunoglobulin chains specifically reactive with predetermined human antigens is problematic, since a reliable source of human antibody-secreting B cells that can evoke an antibody response against predetermined human antigens is needed.

One methodology that can be used to obtain human antibodies that are specifically reactive with human antigens is the production of a transgenic mouse harboring the human immunoglobulin transgene constructs of this invention. Briefly, transgenes containing all or portions of the human immunoglobulin heavy and light chain loci, or transgenes containing synthetic "miniloci" (described *infra*, and in copending applications U.S.S.N. 08/352,322, filed 7 December 1994, U.S.S.N. 07/990,860, filed 16 December 1992, U.S.S.N. 07/810,279 filed 17 December 1991, U.S.S.N. 07/904,068 filed 23 June 1992; U.S.S.N. 07/853,408, filed 18 March 1992, U.S.S.N. 07/574,748 filed August 29, 1990, U.S.S.N. 07/575,962 filed August 31, 1990, and PCT/US91/06185 filed August 28, 1991, each incorporated herein by reference) which comprise essential functional elements of the human heavy and light chain loci, are employed to produce a transgenic nonhuman animal. Such a transgenic nonhuman animal will have the capacity to produce immunoglobulin chains that are encoded by human immunoglobulin genes, and additionally will be capable of making an immune response against human antigens. Thus, such transgenic animals can serve as a source of immune sera reactive with specified human antigens, and B-cells from such transgenic animals can be fused with myeloma cells to produce hybridomas that secrete monoclonal antibodies that are encoded by human immunoglobulin genes and which are specifically reactive with human antigens.

The production of transgenic mice containing various forms of immunoglobulin genes has been reported previously. Rearranged mouse immunoglobulin heavy or light chain genes have been used to produce transgenic mice. In addition, functionally rearranged human Ig genes including the μ or $\gamma 1$ constant region have been expressed in transgenic mice. However, experiments in which the transgene comprises unrearranged (V-D-J or V-J not rearranged) immunoglobulin genes have been variable, in some cases, producing incomplete or minimal rearrangement of the transgene. However, there are no published examples of either rearranged or unrearranged immunoglobulin transgenes which undergo successful isotype switching between C_H genes within a transgene.

The invention also provides a method for identifying candidate hybridomas which secrete a monoclonal antibody comprising a human immunoglobulin chain consisting essentially of a human VDJ sequence in polypeptide linkage to a human constant region sequence. Such candidate hybridomas are identified from a pool of hybridoma clones comprising: (1) hybridoma clones that express immunoglobulin chains consisting essentially of a human VDJ region and a human constant region, and (2) trans-switched hybridomas that express heterohybrid immunoglobulin chains consisting essentially of a human VDJ region and a murine constant region. The supernatant(s) of individual or pooled hybridoma clones is contacted with a predetermined antigen, typically an antigen which is immobilized by adsorption onto a solid substrate (e.g., a microtitre well), under binding conditions to select antibodies having the predetermined antigen binding specificity. An antibody that specifically binds to human constant regions is also contacted with the hybridoma supernatant and predetermined antigen under binding conditions so that the antibody selectively binds to at least one human constant region epitope but substantially does not bind to murine constant region epitopes; thus forming complexes consisting essentially of hybridoma supernatant (transgenic monoclonal antibody) bound to a predetermined antigen and to an antibody that specifically binds human constant regions

(and which may be labeled with a detectable label or reporter). Detection of the formation of such complexes indicates hybridoma clones or pools which express a human immunoglobulin chain.

In a preferred embodiment of the invention, the anti-human constant region immunoglobulin used in screening specifically recognizes a non- μ , non- δ isotype, preferably a α or ϵ , more preferably a γ isotype constant region. Monoclonal antibodies of the γ isotype are preferred (i) because the characteristics of IgG immunoglobulins are preferable to IgM immunoglobulins for some therapeutic applications (e.g., due to the smaller size of the IgG dimers compared to IgM pentamers) and, (ii) because the process of somatic mutation is correlated with the class switch from the μ constant region to the non- μ (e.g., γ) constant regions. Immunoglobulins selected from the population of immunoglobulins that have undergone class switch (e.g., IgG) tend to bind antigen with higher affinity than immunoglobulins selected from the population that has not undergone class switch (e.g., IgM). See, e.g., Lonberg and Huszar. Intern. Rev. Immunol. 13:65-93 (1995) which is incorporated herein by reference.

In one embodiment the candidate hybridomas are first screened for the γ isotype constant region and the pool of IgG-expressing hybridomas is then screened for specific binding to the predetermined antigen.

Thus, according to the method, a transgenic mouse of the invention is immunized with the predetermined antigen to induce an immune response. B cells are collected from the mouse and fused to immortal cells to produce hybridomas. The hybridomas are first screened to identify individual hybridomas secreting Ig of a non- μ , non- δ isotype (e.g., IgG). This set of hybridomas is then screened for specific binding to the predetermined antigen of interest. Screening is carried out using standard techniques as described in, e.g., Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor, New York (1988). Using this method it is

possible to identify high-affinity immunoglobulins (e.g., K_a greater than about 10^7 M^{-1}) practically and efficiently.

Definitions

As used herein, the term "antibody" refers to a glycoprotein comprising at least two light polypeptide chains and two heavy polypeptide chains. Each of the heavy and light polypeptide chains contains a variable region (generally the amino terminal portion of the polypeptide chain) which contains a binding domain which interacts with antigen. Each of the heavy and light polypeptide chains also comprises a constant region of the polypeptide chains (generally the carboxyl terminal portion) which may mediate the binding of the immunoglobulin to host tissues or factors including various cells of the immune system, some phagocytic cells and the first component (C1q) of the classical complement system.

As used herein, a "heterologous antibody" is defined in relation to the transgenic non-human organism producing such an antibody. It is defined as an antibody having an amino acid sequence or an encoding DNA sequence corresponding to that found in an organism not consisting of the transgenic non-human animal, and generally from a species other than that of the transgenic non-human animal.

As used herein, a "heterohybrid antibody" refers to an antibody having a light and heavy chains of different organismal origins. For example, an antibody having a human heavy chain associated with a murine light chain is a heterohybrid antibody.

As used herein, "isotype" refers to the antibody class (e.g., IgM or IgG₁) that is encoded by heavy chain constant region genes.

As used herein, "isotype switching" refers to the phenomenon by which the class, or isotype, of an antibody changes from one Ig class to one of the other Ig classes.

As used herein, "nonswitched isotype" refers to the isotypic class of heavy chain that is produced when no isotype switching has taken place; the C_H gene encoding the nonswitched

isotype is typically the first C_H gene immediately downstream from the functionally rearranged VDJ gene.

As used herein, the term "switch sequence" refers to those DNA sequences responsible for switch recombination. A "switch donor" sequence, typically a μ switch region, will be 5' (i.e., upstream) of the construct region to be deleted during the switch recombination. The "switch acceptor" region will be between the construct region to be deleted and the replacement constant region (e.g., γ , ϵ , etc.). As there is no specific site where recombination always occurs, the final gene sequence will typically not be predictable from the construct.

As used herein, "glycosylation pattern" is defined as the pattern of carbohydrate units that are covalently attached to a protein, more specifically to an immunoglobulin protein. A glycosylation pattern of a heterologous antibody can be characterized as being substantially similar to glycosylation patterns which occur naturally on antibodies produced by the species of the nonhuman transgenic animal, when one of ordinary skill in the art would recognize the glycosylation pattern of the heterologous antibody as being more similar to said pattern of glycosylation in the species of the nonhuman transgenic animal than to the species from which the C_H genes of the transgene were derived.

As used herein, "specific binding" refers to the property of the antibody: (1) to bind to a predetermined antigen with an affinity of at least $1 \times 10^7 \text{ M}^{-1}$, and (2) to preferentially bind to the predetermined antigen with an affinity that is at least two-fold greater than its affinity for binding to a non-specific antigen (e.g., BSA, casein) other than the predetermined antigen or a closely-related antigen.

The term "naturally-occurring" as used herein as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not

been intentionally modified by man in the laboratory is naturally-occurring.

The term "rearranged" as used herein refers to a configuration of a heavy chain or light chain immunoglobulin locus wherein a V segment is positioned immediately adjacent to a D-J or J segment in a conformation encoding essentially a complete V_H or V_L domain, respectively. A rearranged immunoglobulin gene locus can be identified by comparison to germline DNA; a rearranged locus will have at least one recombined heptamer/nonamer homology element.

The term "unrearranged" or "germline configuration" as used herein in reference to a V segment refers to the configuration wherein the V segment is not recombined so as to be immediately adjacent to a D or J segment.

For nucleic acids, the term "substantial homology" indicates that two nucleic acids, or designated sequences thereof, when optimally aligned and compared, are identical, with appropriate nucleotide insertions or deletions, in at least about 80% of the nucleotides, usually at least about 90% to 95%, and more preferably at least about 98 to 99.5% of the nucleotides. Alternatively, substantial homology exists when the segments will hybridize under selective hybridization conditions, to the complement of the strand. The nucleic acids may be present in whole cells, in a cell lysate, or in a partially purified or substantially pure form. A nucleic acid is "isolated" or "rendered substantially pure" when purified away from other cellular components or other contaminants, e.g., other cellular nucleic acids or proteins, by standard techniques, including alkaline/SDS treatment, CsCl banding, column chromatography, agarose gel electrophoresis and others well known in the art. See, F. Ausubel, et al., ed. Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, New York (1987).

The nucleic acid compositions of the present invention, while often in a native sequence (except for modified restriction sites and the like), from either cDNA, genomic or mixtures may be mutated, thereof in accordance with standard techniques to provide gene sequences. For coding

sequences, these mutations, may affect amino acid sequence as desired. In particular, DNA sequences substantially homologous to or derived from native V, D, J, constant, switches and other such sequences described herein are contemplated (where "derived" indicates that a sequence is identical or modified from another sequence).

A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence. With respect to transcription regulatory sequences, operably linked means that the DNA sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame. For switch sequences, operably linked indicates that the sequences are capable of effecting switch recombination.

Transgenic Nonhuman Animals Capable of Producing Heterologous Antibodies

The design of a transgenic non-human animal that responds to foreign antigen stimulation with a heterologous antibody repertoire, requires that the heterologous immunoglobulin transgenes contained within the transgenic animal function correctly throughout the pathway of B-cell development. In a preferred embodiment, correct function of a heterologous heavy chain transgene includes isotype switching. Accordingly, the transgenes of the invention are constructed so as to produce isotype switching and one or more of the following: (1) high level and cell-type specific expression, (2) functional gene rearrangement, (3) activation of and response to allelic exclusion, (4) expression of a sufficient primary repertoire, (5) signal transduction, (6) somatic hypermutation, and (7) domination of the transgene antibody locus during the immune response.

As will be apparent from the following disclosure, not all of the foregoing criteria need be met. For example, in those embodiments wherein the endogenous immunoglobulin loci of the transgenic animal are functionally disrupted, the transgene need not activate allelic exclusion. Further, in

those embodiments wherein the transgene comprises a functionally rearranged heavy and/or light chain immunoglobulin gene, the second criteria of functional gene rearrangement is unnecessary, at least for that transgene which is already rearranged. For background on molecular immunology, see, Fundamental Immunology, 2nd edition (1989), Paul William E., ed. Raven Press, N.Y., which is incorporated herein by reference.

In one aspect of the invention, transgenic non-human animals are provided that contain rearranged, unrearranged or a combination of rearranged and unrearranged heterologous immunoglobulin heavy and light chain transgenes in the germline of the transgenic animal. Each of the heavy chain transgenes comprises at least one C_H gene. In addition, the heavy chain transgene may contain functional isotype switch sequences, which are capable of supporting isotype switching of a heterologous transgene encoding multiple C_H genes in B-cells of the transgenic animal. Such switch sequences may be those which occur naturally in the germline immunoglobulin locus from the species that serves as the source of the transgene C_H genes, or such switch sequences may be derived from those which occur in the species that is to receive the transgene construct (the transgenic animal). For example, a human transgene construct that is used to produce a transgenic mouse may produce a higher frequency of isotype switching events if it incorporates switch sequences similar to those that occur naturally in the mouse heavy chain locus, as presumably the mouse switch sequences are optimized to function with the mouse switch recombinase enzyme system, whereas the human switch sequences are not. Switch sequences made be isolated and cloned by conventional cloning methods, or may be synthesized de novo from overlapping synthetic oligonucleotides designed on the basis of published sequence information relating to immunoglobulin switch region sequences (Mills et al., Nucl. Acids Res. 18:7305-7316 (1991); Sideras et al., Intl. Immunol. 1:631-642 (1989), which are incorporated herein by reference).

For each of the foregoing transgenic animals, functionally rearranged heterologous heavy and light chain immunoglobulin transgenes are found in a significant fraction of the B-cells of the transgenic animal (at least 10 percent).

The transgenes of the invention include a heavy chain transgene comprising DNA encoding at least one variable gene segment, one diversity gene segment, one joining gene segment and at least one constant region gene segment. The immunoglobulin light chain transgene comprises DNA encoding at least one variable gene segment, one joining gene segment and at least one constant region gene segment. The gene segments encoding the light and heavy chain gene segments are heterologous to the transgenic non-human animal in that they are derived from, or correspond to, DNA encoding immunoglobulin heavy and light chain gene segments from a species not consisting of the transgenic non-human animal. In one aspect of the invention, the transgene is constructed such that the individual gene segments are unrearranged, i.e., not rearranged so as to encode a functional immunoglobulin light or heavy chain. Such unrearranged transgenes support recombination of the V, D, and J gene segments (functional rearrangement) and preferably support incorporation of all or a portion of a D region gene segment in the resultant rearranged immunoglobulin heavy chain within the transgenic non-human animal when exposed to antigen.

In an alternate embodiment, the transgenes comprise an unrearranged "mini-locus". Such transgenes typically comprise a substantial portion of the C, D, and J segments as well as a subset of the V gene segments. In such transgene constructs, the various regulatory sequences, e.g. promoters, enhancers, class switch regions, splice-donor and splice-acceptor sequences for RNA processing, recombination signals and the like, comprise corresponding sequences derived from the heterologous DNA. Such regulatory sequences may be incorporated into the transgene from the same or a related species of the non-human animal used in the invention. For example, human immunoglobulin gene segments may be combined in a transgene with a rodent immunoglobulin enhancer sequence for

use in a transgenic mouse. Alternatively, synthetic regulatory sequences may be incorporated into the transgene, wherein such synthetic regulatory sequences are not homologous to a functional DNA sequence that is known to occur naturally in the genomes of mammals. Synthetic regulatory sequences are designed according to consensus rules, such as, for example, those specifying the permissible sequences of a splice-acceptor site or a promoter/enhancer motif. For example, a minilocus comprises a portion of the genomic immunoglobulin locus having at least one internal (i.e., not at a terminus of the portion) deletion of a non-essential DNA portion (e.g., intervening sequence; intron or portion thereof) as compared to the naturally-occurring germline Ig locus.

The invention also includes transgenic animals containing germ line cells having a heavy and light transgene wherein one of the said transgenes contains rearranged gene segments with the other containing unrearranged gene segments. In the preferred embodiments, the rearranged transgene is a light chain immunoglobulin transgene and the unrearranged transgene is a heavy chain immunoglobulin transgene.

The Structure and Generation of Antibodies

The basic structure of all immunoglobulins is based upon a unit consisting of two light polypeptide chains and two heavy polypeptide chains. Each light chain comprises two regions known as the variable light chain region and the constant light chain region. Similarly, the immunoglobulin heavy chain comprises two regions designated the variable heavy chain region and the constant heavy chain region.

The constant region for the heavy or light chain is encoded by genomic sequences referred to as heavy or light constant region gene (C_H) segments. The use of a particular heavy chain gene segment defines the class of immunoglobulin. For example, in humans, the μ constant region gene segments define the IgM class of antibody whereas the use of a γ , γ_2 , γ_3 or γ_4 constant region gene segment defines the IgG class of antibodies as well as the IgG subclasses IgG1 through IgG4.

Similarly, the use of a α_1 or α_2 constant region gene segment defines the IgA class of antibodies as well as the subclasses

IgA1 and IgA2. The δ and ϵ constant region gene segments define the IgD and IgE antibody classes, respectively.

The variable regions of the heavy and light immunoglobulin chains together contain the antigen binding domain of the antibody. Because of the need for diversity in this region of the antibody to permit binding to a wide range of antigens, the DNA encoding the initial or primary repertoire variable region comprises a number of different DNA segments derived from families of specific variable region gene segments. In the case of the light chain variable region, such families comprise variable (V) gene segments and joining (J) gene segments. Thus, the initial variable region of the light chain is encoded by one V gene segment and one J gene segment each selected from the family of V and J gene segments contained in the genomic DNA of the organism. In the case of the heavy chain variable region, the DNA encoding the initial or primary repertoire variable region of the heavy chain comprises one heavy chain V gene segment, one heavy chain diversity (D) gene segment and one J gene segment, each selected from the appropriate V, D and J families of immunoglobulin gene segments in genomic DNA.

In order to increase the diversity of sequences that contribute to forming antibody binding sites, it is preferable that a heavy chain transgene include cis-acting sequences that support functional V-D-J rearrangement that can incorporate all or part of a D region gene sequence in a rearranged V-D-J gene sequence. Typically, at least about 1 percent of expressed transgene-encoded heavy chains (or mRNAs) include recognizable D region sequences in the V region. Preferably, at least about 10 percent of transgene-encoded V regions include recognizable D region sequences, more preferably at least about 30 percent, and most preferably more than 50 percent include recognizable D region sequences.

A recognizable D region sequence is generally at least about eight consecutive nucleotides corresponding to a

sequence present in a D region gene segment of a heavy chain transgene and/or the amino acid sequence encoded by such D region nucleotide sequence. For example, if a transgene includes the D region gene DHQ52, a transgene-encoded mRNA containing the sequence 5'-TAACTGGG-3' located in the V region between a V gene segment sequence and a J gene segment sequence is recognizable as containing a D region sequence, specifically a DHQ52 sequence. Similarly, for example, if a transgene includes the D region gene DHQ52, a transgene-encoded heavy chain polypeptide containing the amino acid sequence -DAF- located in the V region between a V gene segment amino acid sequence and a J gene segment amino acid sequence may be recognizable as containing a D region sequence, specifically a DHQ52 sequence. However, since D region segments may be incorporated in VDJ joining to various extents and in various reading frames, a comparison of the D region area of a heavy chain variable region to the D region segments present in the transgene is necessary to determine the incorporation of particular D segments. Moreover, potential exonuclease digestion during recombination may lead to imprecise V-D and D-J joints during V-D-J recombination.

However, because of somatic mutation and N-region addition, some D region sequences may be recognizable but may not correspond identically to a consecutive D region sequence in the transgene. For example, a nucleotide sequence 5'-CTAAXTGGGG-3' (SEQ ID NO:1), where X is A, T, or G, and which is located in a heavy chain V region and flanked by a V region gene sequence and a J region gene sequence, can be recognized as corresponding to the DHQ52 sequence 5'-CTAACTGGG-3'. Similarly, for example, the polypeptide sequences -DAFDI- (SEQ ID NO:2), -DYFDY- (SEQ ID NO:3, or -GAFDI- (SEQ ID NO:4) located in a V region and flanked on the amino-terminal side by an amino acid sequence encoded by a transgene V gene sequence and flanked on the carboxyterminal side by an amino acid sequence encoded by a transgene J gene sequence is recognizable as a D region sequence.

Therefore, because somatic mutation and N-region addition can produce mutations in sequences derived from a

transgene D region, the following definition is provided as a guide for determining the presence of a recognizable D region sequence. An amino acid sequence or nucleotide sequence is recognizable as a D region sequence if: (1) the sequence is located in a V region and is flanked on one side by a V gene sequence (nucleotide sequence or deduced amino acid sequence) and on the other side by a J gene sequence (nucleotide sequence or deduced amino acid sequence) and (2) the sequence is substantially identical or substantially similar to a known D gene sequence (nucleotide sequence or encoded amino acid sequence).

The term "substantial identity" as used herein denotes a characteristic of a polypeptide sequence or nucleic acid sequence, wherein the polypeptide sequence has at least 50 percent sequence identity compared to a reference sequence, often at least about 80% sequence identity and sometimes more than about 90% sequence identity, and the nucleic acid sequence has at least 70 percent sequence identity compared to a reference sequence. The percentage of sequence identity is calculated excluding small deletions or additions which total less than 35 percent of the reference sequence. The reference sequence may be a subset of a larger sequence, such as an entire D gene; however, the reference sequence is at least 8 nucleotides long in the case of polynucleotides, and at least 3 amino residues long in the case of a polypeptide. Typically, the reference sequence is at least 8 to 12 nucleotides or at least 3 to 4 amino acids, and preferably the reference sequence is 12 to 15 nucleotides or more, or at least 5 amino acids.

The term "substantial similarity" denotes a characteristic of an polypeptide sequence, wherein the polypeptide sequence has at least 80 percent similarity to a reference sequence. The percentage of sequence similarity is calculated by scoring identical amino acids or positional conservative amino acid substitutions as similar. A positional conservative amino acid substitution is one that can result from a single nucleotide substitution; a first amino acid is replaced by a second amino acid where a codon

for the first amino acid and a codon for the second amino acid can differ by a single nucleotide substitution. Thus, for example, the sequence -Lys-Glu-Arg-Val- (SEQ ID NO:5) is substantially similar to the sequence -Asn-Asp-Ser-Val- (SEQ ID NO:6), since the codon sequence -AAA-GAA-AGA-GUU- (SEQ ID NO:7) can be mutated to -AAC-GAC-AGC-GUU- (SEQ ID NO:8) by introducing only 3 substitution mutations, single nucleotide substitutions in three of the four original codons. The reference sequence may be a subset of a larger sequence, such as an entire D gene; however, the reference sequence is at least 4 amino residues long. Typically, the reference sequence is at least 5 amino acids, and preferably the reference sequence is 6 amino acids or more.

The Primary Repertoire

The process for generating DNA encoding the heavy and light chain immunoglobulin genes occurs primarily in developing B-cells. Prior to the joining of various immunoglobulin gene segments, the V, D, J and constant (C) gene segments are found, for the most part, in clusters of V, D, J and C gene segments in the precursors of primary repertoire B-cells. Generally, all of the gene segments for a heavy or light chain are located in relatively close proximity on a single chromosome. Such genomic DNA prior to recombination of the various immunoglobulin gene segments is referred to herein as "unrearranged" genomic DNA. During B-cell differentiation, one of each of the appropriate family members of the V, D, J (or only V and J in the case of light chain genes) gene segments are recombined to form functionally rearranged heavy and light immunoglobulin genes. Such functional rearrangement is of the variable region segments to form DNA encoding a functional variable region. This gene segment rearrangement process appears to be sequential. First, heavy chain D-to-J joints are made, followed by heavy chain V-to-DJ joints and light chain V-to-J joints. The DNA encoding this initial form of a functional variable region in a light and/or heavy chain is referred to as "functionally rearranged DNA" or "rearranged DNA". In the case of the heavy

chain, such DNA is referred to as "rearranged heavy chain DNA" and in the case of the light chain, such DNA is referred to as "rearranged light chain DNA". Similar language is used to describe the functional rearrangement of the transgenes of the invention.

The recombination of variable region gene segments to form functional heavy and light chain variable regions is mediated by recombination signal sequences (RSS's) that flank recombinationally competent V, D and J segments. RSS's necessary and sufficient to direct recombination, comprise a dyad-symmetric heptamer, an AT-rich nonamer and an intervening spacer region of either 12 or 23 base pairs. These signals are conserved among the different loci and species that carry out D-J (or V-J) recombination and are functionally interchangeable. See Oettinger, et al. (1990), Science, 248, 1517-1523 and references cited therein. The heptamer comprises the sequence CACAGTG or its analogue followed by a spacer of unconserved sequence and then a nonamer having the sequence ACAAAAACC or its analogue. These sequences are found on the J, or downstream side, of each V and D gene segment. Immediately preceding the germline D and J segments are again two recombination signal sequences, first the nonamer and then the heptamer again separated by an unconserved sequence. The heptameric and nonameric sequences following a V_L , V_H or D segment are complementary to those preceding the J_L , D or J_H segments with which they recombine. The spacers between the heptameric and nonameric sequences are either 12 base pairs long or between 22 and 24 base pairs long.

In addition to the rearrangement of V, D and J segments, further diversity is generated in the primary repertoire of immunoglobulin heavy and light chain by way of variable recombination between the V and J segments in the light chain and between the D and J segments of the heavy chain. Such variable recombination is generated by variation in the exact place at which such segments are joined. Such variation in the light chain typically occurs within the last codon of the V gene segment and the first codon of the J segment. Similar imprecision in joining occurs on the heavy

chain chromosome between the D and J_H segments and may extend over as many as 10 nucleotides. Furthermore, several nucleotides may be inserted between the D and J_H and between the V_H and D gene segments which are not encoded by genomic DNA. The addition of these nucleotides is known as N-region diversity.

After VJ and/or VDJ rearrangement, transcription of the rearranged variable region and one or more constant region gene segments located downstream from the rearranged variable region produces a primary RNA transcript which upon appropriate RNA splicing results in an mRNA which encodes a full length heavy or light immunoglobulin chain. Such heavy and light chains include a leader signal sequence to effect secretion through and/or insertion of the immunoglobulin into the transmembrane region of the B-cell. The DNA encoding this signal sequence is contained within the first exon of the V segment used to form the variable region of the heavy or light immunoglobulin chain. Appropriate regulatory sequences are also present in the mRNA to control translation of the mRNA to produce the encoded heavy and light immunoglobulin polypeptides which upon proper association with each other form an antibody molecule.

The net effect of such rearrangements in the variable region gene segments and the variable recombination which may occur during such joining, is the production of a primary antibody repertoire. Generally, each B-cell which has differentiated to this stage, produces a single primary repertoire antibody. During this differentiation process, cellular events occur which suppress the functional rearrangement of gene segments other than those contained within the functionally rearranged Ig gene. The process by which diploid B-cells maintain such mono-specificity is termed allelic exclusion.

The Secondary Repertoire

B-cell clones expressing immunoglobulins from within the set of sequences comprising the primary repertoire are immediately available to respond to foreign antigens. Because

of the limited diversity generated by simple VJ and VDJ joining, the antibodies produced by the so-called primary response are of relatively low affinity. Two different types of B-cells make up this initial response: precursors of primary antibody-forming cells and precursors of secondary repertoire B-cells (Linton et al., Cell 59:1049-1059 (1989)). The first type of B-cell matures into IgM-secreting plasma cells in response to certain antigens. The other B-cells respond to initial exposure to antigen by entering a T-cell dependent maturation pathway.

During the T-cell dependent maturation of antigen stimulated B-cell clones, the structure of the antibody molecule on the cell surface changes in two ways: the constant region switches to a non-IgM subtype and the sequence of the variable region can be modified by multiple single amino acid substitutions to produce a higher affinity antibody molecule.

As previously indicated, each variable region of a heavy or light Ig chain contains an antigen binding domain. It has been determined by amino acid and nucleic acid sequencing that somatic mutation during the secondary response occurs throughout the V region including the three complementary determining regions (CDR1, CDR2 and CDR3) also referred to as hypervariable regions 1, 2 and 3 (Kabat et al. Sequences of Proteins of Immunological Interest (1991) U.S. Department of Health and Human Services, Washington, DC, incorporated herein by reference. The CDR1 and CDR2 are located within the variable gene segment whereas the CDR3 is largely the result of recombination between V and J gene segments or V, D and J gene segments. Those portions of the variable region which do not consist of CDR1, 2 or 3 are commonly referred to as framework regions designated FR1, FR2, FR3 and FR4. See Fig. 1. During hypermutation, the rearranged DNA is mutated to give rise to new clones with altered Ig molecules. Those clones with higher affinities for the foreign antigen are selectively expanded by helper T-cells, giving rise to affinity maturation of the expressed antibody. Clonal selection typically results in expression of clones containing new mutation within the CDR1, 2 and/or 3

regions. However, mutations outside these regions also occur which influence the specificity and affinity of the antigen binding domain.

Transgenic Non-Human Animals Capable of Producing Heterologous Antibody

Transgenic non-human animals in one aspect of the invention are produced by introducing at least one of the immunoglobulin transgenes of the invention (discussed hereinafter) into a zygote or early embryo of a non-human animal. The non-human animals which are used in the invention generally comprise any mammal which is capable of rearranging immunoglobulin gene segments to produce a primary antibody response. Such nonhuman transgenic animals may include, for example, transgenic pigs, transgenic rats, transgenic rabbits, transgenic cattle, and other transgenic animal species, particularly mammalian species, known in the art. A particularly preferred non-human animal is the mouse or other members of the rodent family.

However, the invention is not limited to the use of mice. Rather, any non-human mammal which is capable of mounting a primary and secondary antibody response may be used. Such animals include non-human primates, such as chimpanzee, bovine, ovine, and porcine species, other members of the rodent family, e.g. rat, as well as rabbit and guinea pig. Particular preferred animals are mouse, rat, rabbit and guinea pig, most preferably mouse.

In one embodiment of the invention, various gene segments from the human genome are used in heavy and light chain transgenes in an unrearranged form. In this embodiment, such transgenes are introduced into mice. The unrearranged gene segments of the light and/or heavy chain transgene have DNA sequences unique to the human species which are distinguishable from the endogenous immunoglobulin gene segments in the mouse genome. They may be readily detected in unrearranged form in the germ line and somatic cells not consisting of B-cells and in rearranged form in B-cells.

In an alternate embodiment of the invention, the transgenes comprise rearranged heavy and/or light immunoglobulin transgenes. Specific segments of such transgenes corresponding to functionally rearranged VDJ or VJ segments, contain immunoglobulin DNA sequences which are also clearly distinguishable from the endogenous immunoglobulin gene segments in the mouse.

Such differences in DNA sequence are also reflected in the amino acid sequence encoded by such human immunoglobulin transgenes as compared to those encoded by mouse B-cells. Thus, human immunoglobulin amino acid sequences may be detected in the transgenic non-human animals of the invention with antibodies specific for immunoglobulin epitopes encoded by human immunoglobulin gene segments.

Transgenic B-cells containing unrearranged transgenes from human or other species functionally recombine the appropriate gene segments to form functionally rearranged light and heavy chain variable regions. It will be readily apparent that the antibody encoded by such rearranged transgenes has a DNA and/or amino acid sequence which is heterologous to that normally encountered in the nonhuman animal used to practice the invention.

Unrearranged Transgenes

As used herein, an "unrearranged immunoglobulin heavy chain transgene" comprises DNA encoding at least one variable gene segment, one diversity gene segment, one joining gene segment and one constant region gene segment. Each of the gene segments of said heavy chain transgene are derived from, or has a sequence corresponding to, DNA encoding immunoglobulin heavy chain gene segments from a species not consisting of the non-human animal into which said transgene is introduced. Similarly, as used herein, an "unrearranged immunoglobulin light chain transgene" comprises DNA encoding at least one variable gene segment, one joining gene segment and at least one constant region gene segment wherein each gene segment of said light chain transgene is derived from, or has a sequence corresponding to, DNA encoding immunoglobulin

light chain gene segments from a species not consisting of the non-human animal into which said light chain transgene is introduced.

Such heavy and light chain transgenes in this aspect of the invention contain the above-identified gene segments in an unrearranged form. Thus, interposed between the V, D and J segments in the heavy chain transgene and between the V and J segments on the light chain transgene are appropriate recombination signal sequences (RSS's). In addition, such transgenes also include appropriate RNA splicing signals to join a constant region gene segment with the VJ or VDJ rearranged variable region.

In order to facilitate isotype switching within a heavy chain transgene containing more than one C region gene segment, e.g. C μ and C γ 1 from the human genome, as explained below "switch regions" are incorporated upstream from each of the constant region gene segments and downstream from the variable region gene segments to permit recombination between such constant regions to allow for immunoglobulin class switching, e.g. from IgM to IgG. Such heavy and light immunoglobulin transgenes also contain transcription control sequences including promoter regions situated upstream from the variable region gene segments which typically contain TATA motifs. A promoter region can be defined approximately as a DNA sequence that, when operably linked to a downstream sequence, can produce transcription of the downstream sequence. Promoters may require the presence of additional linked cis-acting sequences in order to produce efficient transcription. In addition, other sequences that participate in the transcription of sterile transcripts are preferably included. Examples of sequences that participate in expression of sterile transcripts can be found in the published literature, including Rothman et al., Intl. Immunol. 2:621-627 (1990); Reid et al., Proc. Natl. Acad. Sci. USA 86:840-844 (1989); Stavnezer et al., Proc. Natl. Acad. Sci. USA 85:7704-7708 (1988); and Mills et al., Nucl. Acids Res. 18:7305-7316 (1991), each of which is incorporated herein by reference. These sequences typically include about at least

50 bp immediately upstream of a switch region, preferably about at least 200 bp upstream of a switch region; and more preferably about at least 200-1000 bp or more upstream of a switch region. Suitable sequences occur immediately upstream of the human S_{Y1} , S_{Y2} , S_{Y3} , S_{Y4} , $S_{\alpha 1}$, $S_{\alpha 2}$, and S_{ϵ} switch regions; the sequences immediately upstream of the human S_{Y1} , and S_{Y3} switch regions can be used to advantage, with S_{Y1} generally preferred. Alternatively, or in combination, murine Ig switch sequences may be used; it may frequently be advantageous to employ Ig switch sequences of the same species as the transgenic non-human animal. Furthermore, interferon (IFN) inducible transcriptional regulatory elements, such as IFN-inducible enhancers, are preferably included immediately upstream of transgene switch sequences.

In addition to promoters, other regulatory sequences which function primarily in B-lineage cells are used. Thus, for example, a light chain enhancer sequence situated preferably between the J and constant region gene segments on the light chain transgene is used to enhance transgene expression, thereby facilitating allelic exclusion. In the case of the heavy chain transgene, regulatory enhancers are also employed. Such regulatory sequences are used to maximize the transcription and translation of the transgene so as to induce allelic exclusion and to provide relatively high levels of transgene expression.

Although the foregoing promoter and enhancer regulatory control sequences have been generically described, such regulatory sequences may be heterologous to the nonhuman animal being derived from the genomic DNA from which the heterologous transgene immunoglobulin gene segments are obtained. Alternately, such regulatory gene segments are derived from the corresponding regulatory sequences in the genome of the non-human animal, or closely related species, which contains the heavy and light transgene.

In the preferred embodiments, gene segments are derived from human beings. The transgenic non-human animals harboring such heavy and light transgenes are capable of mounting an Ig-mediated immune response to a specific antigen

administered to such an animal. B-cells are produced within such an animal which are capable of producing heterologous human antibody. After immortalization, and the selection for an appropriate monoclonal antibody (Mab), e.g. a hybridoma, a source of therapeutic human monoclonal antibody is provided. Such human Mabs have significantly reduced immunogenicity when therapeutically administered to humans.

Although the preferred embodiments disclose the construction of heavy and light transgenes containing human gene segments, the invention is not so limited. In this regard, it is to be understood that the teachings described herein may be readily adapted to utilize immunoglobulin gene segments from a species other than human beings. For example, in addition to the therapeutic treatment of humans with the antibodies of the invention, therapeutic antibodies encoded by appropriate gene segments may be utilized to generate monoclonal antibodies for use in the veterinary sciences.

Rearranged Transgenes

In an alternative embodiment, transgenic nonhuman animals contain functionally at least one rearranged heterologous heavy chain immunoglobulin transgene in the germline of the transgenic animal. Such animals contain primary repertoire B-cells that express such rearranged heavy transgenes. Such B-cells preferably are capable of undergoing somatic mutation when contacted with an antigen to form a heterologous antibody having high affinity and specificity for the antigen. Said rearranged transgenes will contain at least two C_H genes and the associated sequences required for isotype switching.

The invention also includes transgenic animals containing germ line cells having heavy and light transgenes wherein one of the said transgenes contains rearranged gene segments with the other containing unrearranged gene segments. In such animals, the heavy chain transgenes shall have at least two C_H genes and the associated sequences required for isotype switching.

The invention further includes methods for generating a synthetic variable region gene segment repertoire to be used in the transgenes of the invention. The method comprises generating a population of immunoglobulin V segment DNAs wherein each of the V segment DNAs encodes an immunoglobulin V segment and contains at each end a cleavage recognition site of a restriction endonuclease. The population of immunoglobulin V segment DNAs is thereafter concatenated to form the synthetic immunoglobulin V segment repertoire. Such synthetic variable region heavy chain transgenes shall have at least two C_H genes and the associated sequences required for isotype switching.

Isotype Switching

In the development of a B lymphocyte, the cell initially produces IgM with a binding specificity determined by the productively rearranged V_H and V_L regions. Subsequently, each B cell and its progeny cells synthesize antibodies with the same L and H chain V regions, but they may switch the isotype of the H chain.

The use of μ or δ constant regions is largely determined by alternate splicing, permitting IgM and IgD to be coexpressed in a single cell. The other heavy chain isotypes (γ , α , and ϵ) are only expressed natively after a gene rearrangement event deletes the C_μ and C_δ exons. This gene rearrangement process, termed isotype switching, typically occurs by recombination between so called switch segments located immediately upstream of each heavy chain gene (except δ). The individual switch segments are between 2 and 10 kb in length, and consist primarily of short repeated sequences. The exact point of recombination differs for individual class switching events. Investigations which have used solution hybridization kinetics or Southern blotting with cDNA-derived C_H probes have confirmed that switching can be associated with loss of C_H sequences from the cell.

The switch (S) region of the μ gene, S_μ , is located about 1 to 2 kb 5' to the coding sequence and is composed of numerous tandem repeats of sequences of the form

(GAGCT)_n(GGGGT) (SEQ ID NOS:9-24), where n is usually 2 to 5 but can range as high as 17. (See T. Nikaido et al. Nature 292:845-848 (1981)).

Similar internally repetitive switch sequences spanning several kilobases have been found 5' of the other C_H genes. The S_α region has been sequenced and found to consist of tandemly repeated 80-bp homology units, whereas murine S_{Y2a}, S_{Y2b}, and S_{Y3} all contain repeated 49-bp homology units very similar to each other. (See, P. Szurek et al., J. Immunol 135:620-626 (1985) and T. Nikaido et al., J. Biol. Chem. 257:7322-7329 (1982), which are incorporated herein by reference.) All the sequenced S regions include numerous occurrences of the pentamers GAGCT and GGGGT that are the basic repeated elements of the S_μ gene (T. Nikaido et al., J. Biol. Chem. 257:7322-7329 (1982) which is incorporated herein by reference); in the other S regions these pentamers are not precisely tandemly repeated as in S_μ, but instead are embedded in larger repeat units. The S_{Y1} region has an additional higher-order structure: two direct repeat sequences flank each of two clusters of 49-bp tandem repeats. (See M. R. Mowatt et al., J. Immunol. 136:2674-2683 (1986), which is incorporated herein by reference).

Switch regions of human H chain genes have been found to be very similar to their mouse homologs. Indeed, similarity between pairs of human and mouse clones 5' to the C_H genes has been found to be confined to the S regions, a fact that confirms the biological significance of these regions.

A switch recombination between μ and α genes produces a composite S_μ-S_α sequence. Typically, there is no specific site, either in S_μ or in any other S region, where the recombination always occurs.

Generally, unlike the enzymatic machinery of V-J recombination, the switch machinery can apparently accommodate different alignments of the repeated homologous regions of germline S precursors and then join the sequences at different positions within the alignment. (See, T. H. Rabbits et al., Nucleic Acids Res. 9:4509-4524 (1981) and J. Ravetch et al.,

Proc. Natl. Acad. Sci. USA 77:6734-6738 (1980), which are incorporated herein by reference.)

The exact details of the mechanism(s) of selective activation of switching to a particular isotype are unknown. Although exogenous influences such as lymphokines and cytokines might upregulate isotype-specific recombinases, it is also possible that the same enzymatic machinery catalyzes switches to all isotypes and that specificity lies in targeting this machinery to specific switch regions.

The T-cell-derived lymphokines IL-4 and IFN γ have been shown to specifically promote the expression of certain isotypes: in the mouse, IL-4 decreases IgM, IgG2a, IgG2b, and IgG3 expression and increases IgE and IgG1 expression; while IFN γ selectively stimulates IgG2a expression and antagonizes the IL-4-induced increase in IgE and IgG1 expression (Coffman et al., J. Immunol. 136: 949 (1986) and Snapper et al., Science 236: 944 (1987), which are incorporated herein by reference). A combination of IL-4 and IL-5 promotes IgA expression (Coffman et al., J. Immunol. 139: 3685 (1987), which is incorporated herein by reference).

Most of the experiments implicating T-cell effects on switching have not ruled out the possibility that the observed increase in cells with particular switch recombinations might reflect selection of preswitched or precommitted cells; but the most likely explanation is that the lymphokines actually promote switch recombination.

Induction of class switching appears to be associated with sterile transcripts that initiate upstream of the switch segments (Lutzker et al., Mol. Cell. Biol. 8:1849 (1988); Stavnezer et al., Proc. Natl. Acad. Sci. USA 85:7704 (1988); Esser and Radbruch, EMBO J. 8:483 (1989); Berton et al., Proc. Natl. Acad. Sci. USA 86:2829 (1989); Rothman et al., Int. Immunol. 2:621 (1990), each of which is incorporated by reference). For example, the observed induction of the γ 1 sterile transcript by IL-4 and inhibition by IFN- γ correlates with the observation that IL-4 promotes class switching to γ 1 in B-cells in culture, while IFN- γ inhibits γ 1 expression. Therefore, the inclusion of regulatory sequences that affect

the transcription of sterile transcripts may also affect the rate of isotype switching. For example, increasing the transcription of a particular sterile transcript typically can be expected to enhance the frequency of isotype switch recombination involving adjacent switch sequences.

For these reasons, it is preferable that transgenes incorporate transcriptional regulatory sequences within about 1-2 kb upstream of each switch region that is to be utilized for isotype switching. These transcriptional regulatory sequences preferably include a promoter and an enhancer element, and more preferably include the 5' flanking (i.e., upstream) region that is naturally associated (i.e., occurs in germline configuration) with a switch region. This 5' flanking region is typically about at least 50 nucleotides in length, preferably about at least 200 nucleotides in length, and more preferably at least 500-1000 nucleotides.

Although a 5' flanking sequence from one switch region can be operably linked to a different switch region for transgene construction (e.g., a 5' flanking sequence from the human $S_{\gamma 1}$ switch can be grafted immediately upstream of the $S_{\alpha 1}$ switch; a murine $S_{\gamma 1}$ flanking region can be grafted adjacent to a human $\gamma 1$ switch sequence; or the murine $S_{\gamma 1}$ switch can be grafted onto the human $\gamma 1$ coding region), in some embodiments it is preferred that each switch region incorporated in the transgene construct have the 5' flanking region that occurs immediately upstream in the naturally occurring germline configuration.

Monoclonal Antibodies

Monoclonal antibodies can be obtained by various techniques familiar to those skilled in the art. Briefly, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (see, Kohler and Milstein, Eur. J. Immunol., 6:511-519 (1976)). Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies

of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Various techniques useful in these arts are discussed, for example, in Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor, New York (1988) including: immunization of animals to produce immunoglobulins; production of monoclonal antibodies; labeling immunoglobulins for use as probes; immunoaffinity purification; and immunoassays.

The Transgenic Primary Repertoire

A. The Human Immunoglobulin Loci

An important requirement for transgene function is the generation of a primary antibody repertoire that is diverse enough to trigger a secondary immune response for a wide range of antigens. The rearranged heavy chain gene consists of a signal peptide exon, a variable region exon and a tandem array of multi-domain constant region regions, each of which is encoded by several exons. Each of the constant region genes encode the constant portion of a different class of immunoglobulins. During B-cell development, V region proximal constant regions are deleted leading to the expression of new heavy chain classes. For each heavy chain class, alternative patterns of RNA splicing give rise to both transmembrane and secreted immunoglobulins.

The human heavy chain locus is estimated to consist of approximately 200 V gene segments (current data supports the existence of about 50-100 V gene segments) spanning 2 Mb, approximately 30 D gene segments spanning about 40 kb, six J segments clustered within a 3 kb span, and nine constant region gene segments spread out over approximately 300 kb. The entire locus spans approximately 2.5 Mb of the distal portion of the long arm of chromosome 14.

B. Gene Fragment Transgenes

1. Heavy Chain Transgene

In a preferred embodiment, immunoglobulin heavy and light chain transgenes comprise unrearranged genomic DNA from humans. In the case of the heavy chain, a preferred transgene comprises a NotI fragment having a length between 670 to 830 kb. The length of this fragment is ambiguous because the 3' restriction site has not been accurately mapped. It is known, however, to reside between the $\alpha 1$ and $\psi\alpha$ gene segments. This fragment contains members of all six of the known V_H families, the D and J gene segments, as well as the μ , δ , $\gamma 3$, $\gamma 1$ and $\alpha 1$ constant regions (Berman et al., EMBO J. 7:727-738 (1988), which is incorporated herein by reference). A transgenic mouse line containing this transgene correctly expresses a heavy chain class required for B-cell development (IgM) and at least one switched heavy chain class (IgG₁), in conjunction with a sufficiently large repertoire of variable regions to trigger a secondary response for most antigens.

2. Light Chain Transgene

A genomic fragment containing all of the necessary gene segments and regulatory sequences from a human light chain locus may be similarly constructed. Such transgenes are constructed as described in the Examples and in copending application, entitled "Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies," filed August 29, 1990, under U.S.S.N. 07/574,748.

C. Transgenes Generated Intracellularly by In Vivo Recombination

It is not necessary to isolate the all or part of the heavy chain locus on a single DNA fragment. Thus, for example, the 670-830 kb NotI fragment from the human immunoglobulin heavy chain locus may be formed in vivo in the non-human animal during transgenesis. Such in vivo transgene construction is produced by introducing two or more overlapping DNA fragments into an embryonic nucleus of the non-human animal. The overlapping portions of the DNA fragments have DNA sequences which are substantially homologous. Upon exposure to the recombinases contained

within the embryonic nucleus, the overlapping DNA fragments homologously recombined in proper orientation to form the 670-830 kb NotI heavy chain fragment.

In vivo transgene construction can be used to form any number of immunoglobulin transgenes which because of their size are otherwise difficult, or impossible, to make or manipulate by present technology. Thus, in vivo transgene construction is useful to generate immunoglobulin transgenes which are larger than DNA fragments which may be manipulated by YAC vectors (Murray and Szostak, Nature 305:189-193 (1983)). Such in vivo transgene construction may be used to introduce into a non-human animal substantially the entire immunoglobulin loci from a species not consisting of the transgenic non-human animal.

In addition to forming genomic immunoglobulin transgenes, in vivo homologous recombination may also be utilized to form "mini-locus" transgenes as described in the Examples.

In the preferred embodiments utilizing in vivo transgene construction, each overlapping DNA fragment preferably has an overlapping substantially homologous DNA sequence between the end portion of one DNA fragment and the end portion of a second DNA fragment. Such overlapping portions of the DNA fragments preferably comprise about 500 bp to about 2000 bp, most preferably 1.0 kb to 2.0 kb. Homologous recombination of overlapping DNA fragments to form transgenes in vivo is further described in commonly assigned U.S. Patent Application entitled "Intracellular Generation of DNA by Homologous Recombination of DNA Fragments" filed August 29, 1990, under U.S.S.N. 07/574,747.

D. Minilocus Transgenes

As used herein, the term "immunoglobulin minilocus" refers to a DNA sequence (which may be within a longer sequence), usually of less than about 150 kb, typically between about 25 and 100 kb, containing at least one each of the following: a functional variable (V) gene segment, a functional joining (J) region segment, at least one functional constant (C) region gene segment, and--if it is a heavy chain minilocus--a functional diversity (D) region segment, such that said DNA sequence contains at least one substantial discontinuity (e.g., a deletion, usually of at least about 2 to 5 kb, preferably 10-25 kb or more, relative to the homologous genomic DNA sequence). A light chain minilocus transgene will be at least 25 kb in length, typically 50 to 60 kb. A heavy chain transgene will typically be about 70 to 80 kb in length, preferably at least about 60 kb with two constant regions operably linked to switch regions. Furthermore, the individual elements of the minilocus are preferably in the germline configuration and capable of undergoing gene rearrangement in the pre-B cell of a transgenic animal so as to express functional antibody molecules with diverse antigen specificities encoded entirely by the elements of the minilocus. Further, a heavy chain minilocus comprising at least two C_H genes and the requisite switching sequences is typically capable of undergoing isotype switching, so that functional antibody molecules of different immunoglobulin classes will be generated. Such isotype switching may occur in vivo in B-cells residing within the transgenic nonhuman animal, or may occur in cultured cells of the B-cell lineage which have been explanted from the transgenic nonhuman animal.

In an alternate preferred embodiment, immunoglobulin heavy chain transgenes comprise one or more of each of the V_H , D, and J_H gene segments and two or more of the C_H genes. At least one of each appropriate type gene segment is incorporated into the minilocus transgene. With regard to the C_H segments for the heavy chain transgene, it is preferred that

the transgene contain at least one μ gene segment and at least one other constant region gene segment, more preferably a γ gene segment, and most preferably $\gamma 3$ or $\gamma 1$. This preference is to allow for class switching between IgM and IgG forms of the encoded immunoglobulin and the production of a secretable form of high affinity non-IgM immunoglobulin. Other constant region gene segments may also be used such as those which encode for the production of IgD, IgA and IgE.

Those skilled in the art will also construct transgenes wherein the order of occurrence of heavy chain C_H genes will be different from the naturally-occurring spatial order found in the germline of the species serving as the donor of the C_H genes.

Additionally, those skilled in the art can select C_H genes from more than one individual of a species (e.g., allogeneic C_H genes) and incorporate said genes in the transgene as supernumerary C_H genes capable of undergoing isotype switching; the resultant transgenic nonhuman animal may then, in some embodiments, make antibodies of various classes including all of the allotypes represented in the species from which the transgene C_H genes were obtained.

Still further, those skilled in the art can select C_H genes from different species to incorporate into the transgene. Functional switch sequences are included with each C_H gene, although the switch sequences used are not necessarily those which occur naturally adjacent to the C_H gene. Interspecies C_H gene combinations will produce a transgenic nonhuman animal which may produce antibodies of various classes corresponding to C_H genes from various species. Transgenic nonhuman animals containing interspecies C_H transgenes may serve as the source of B-cells for constructing hybridomas to produce monoclonals for veterinary uses.

The heavy chain J region segments in the human comprise six functional J segments and three pseudo genes clustered in a 3 kb stretch of DNA. Given its relatively compact size and the ability to isolate these segments together with the μ gene and the 5' portion of the δ gene on a single 23 kb SfiI/SpeI fragment (Sado et al., Biochem.

Biophys. Res. Comm. 154:264271 (1988), which is incorporated herein by reference), it is preferred that all of the J region gene segments be used in the mini-locus construct. Since this fragment spans the region between the μ and δ genes, it is likely to contain all of the 3' cis-linked regulatory elements required for μ expression. Furthermore, because this fragment includes the entire J region, it contains the heavy chain enhancer and the μ switch region (Mills et al., Nature 306:809 (1983); Yancopoulos and Alt, Ann. Rev. Immunol. 4:339-368 (1986), which are incorporated herein by reference). It also contains the transcription start sites which trigger VDJ joining to form primary repertoire B-cells (Yancopoulos and Alt, Cell 40:271-281 (1985), which is incorporated herein by reference). Alternatively, a 36 kb BssHII/SpeI fragment, which includes part on the D region, may be used in place of the 23 kb SfiI/SpeI fragment. The use of such a fragment increases the amount of 5' flanking sequence to facilitate efficient D-to-J joining.

The human D region consists of 4 homologous 9 kb subregions, linked in tandem (Siebenlist, et al. (1981), Nature, 294, 631-635). Each subregion contains up to 10 individual D segments. Some of these segments have been mapped and are shown in Fig. 4. Two different strategies are used to generate a mini-locus D region. The first strategy involves using only those D segments located in a short contiguous stretch of DNA that includes one or two of the repeated D subregions. A candidate is a single 15 kb fragment that contains 12 individual D segments. This piece of DNA consists of 2 contiguous EcoRI fragments and has been completely sequenced (Ichihara, et al. (1988), EMBO J., 7, 4141-4150). Twelve D segments should be sufficient for a primary repertoire. However, given the dispersed nature of the D region, an alternative strategy is to ligate together several non-contiguous D-segment containing fragments, to produce a smaller piece of DNA with a greater number of segments. Additional D-segment genes can be identified, for example, by the presence of characteristic flanking nonamer

and heptamer sequences, supra, and by reference to the literature.

At least one, and preferably more than one V gene segment is used to construct the heavy chain minilocus transgene. Rearranged or unrearranged V segments with or without flanking sequences can be isolated as described in copending applications, U.S.S.N. 07/574,748 filed August 29, 1990, PCT/US91/06185 filed August 28, 1991, and U.S.S.N. 07/810,279 filed December 17, 1991, each of which is incorporated herein by reference.

Rearranged or unrearranged V segments, D segments, J segments, and C genes, with or without flanking sequences, can be isolated as described in copending applications U.S.S.N. 07/574,748 filed August 29, 1990 and PCT/US91/06185 filed August 28, 1991.

A minilocus light chain transgene may be similarly constructed from the human λ or κ immunoglobulin locus.

Thus, for example, an immunoglobulin heavy chain minilocus transgene construct, e.g., of about 75 kb, encoding V, D, J and constant region sequences can be formed from a plurality of DNA fragments, with each sequence being substantially homologous to human gene sequences. Preferably, the sequences are operably linked to transcription regulatory sequences and are capable of undergoing rearrangement. With two or more appropriately placed constant region sequences (e.g., μ and γ) and switch regions, switch recombination also occurs. An exemplary light chain transgene construct can be formed similarly from a plurality of DNA fragments, substantially homologous to human DNA and capable of undergoing rearrangement, as described in copending application, U.S.S.N. 07/574,748 filed August 29, 1990.

E. Transgene Constructs Capable of Isotype Switching

Ideally, transgene constructs that are intended to undergo class switching should include all of the cis-acting sequences necessary to regulate sterile transcripts. Naturally occurring switch regions and upstream promoters and regulatory sequences (e.g., IFN-inducible elements) are

preferred cis-acting sequences that are included in transgene constructs capable of isotype switching. About at least 50 basepairs, preferably about at least 200 basepairs, and more preferably at least 500 to 1000 basepairs or more of sequence immediately upstream of a switch region, preferably a human γ_1 switch region, should be operably linked to a switch sequence, preferably a human γ_1 switch sequence. Further, switch regions can be linked upstream of (and adjacent to) C_H genes that do not naturally occur next to the particular switch region. For example, but not for limitation, a human γ_1 switch region may be linked upstream from a human $\alpha_2 C_H$ gene, or a murine γ_1 switch may be linked to a human C_H gene.

An alternative method for obtaining non-classical isotype switching (e.g., δ -associated deletion) in transgenic mice involves the inclusion of the 400 bp direct repeat sequences ($\sigma\mu$ and $\epsilon\mu$) that flank the human μ gene (Yasui et al., Eur. J. Immunol. 19:1399 (1989)). Homologous recombination between these two sequences deletes the μ gene in IgD-only B-cells. Heavy chain transgenes can be represented by the following formulaic description:

$$(V_H)_x - (D)_y - (J_H)_z - (S_D)_m - (C_1)_n - [(T) - (S_A)_p - (C_2)]_q$$

where:

V_H is a heavy chain variable region gene segment,

D is a heavy chain D (diversity) region gene segment,

J_H is a heavy chain J (joining) region gene segment,

S_D is a donor region segment capable of participating in a recombination event with the S_A acceptor region segments such that isotype switching occurs,

C_1 is a heavy chain constant region gene segment encoding an isotype utilized in for B cell development (e.g., μ or δ),

T is a cis-acting transcriptional regulatory region segment containing at least a promoter,

S_A is an acceptor region segment capable of participating in a recombination event with selected S_D donor region segments, such that isotype switching occurs,

C_2 is a heavy chain constant region gene segment encoding

60D

an isotype other than μ (e.g., γ_1 , γ_2 , γ_3 , γ_4 , α_1 , α_2 , ϵ). x , y , z , m , n , p , and q are integers. x is 1-100, n is 0-10, y is 1-50, p is 1-10, z is 1-50, q is 0-50, m is 0-10. Typically, when the transgene is capable of isotype switching, q must be at least 1, m is at least 1, n is at least 1, and m is greater than or equal to n .

V_H , D , J_H , S_D , C_1 , T , S_A , and C_2 segments may be selected from various species, preferably mammalian species, and more preferably from human and murine germline DNA.

V_H segments may be selected from various species, but are preferably selected from V_H segments that occur naturally in the human germline, such as V_{H251} . Typically about 2 V_H gene segments are included, preferably about 4 V_H segments are included, and most preferably at least about 10 V_H segments are included.

At least one D segment is typically included, although at least 10 D segments are preferably included, and some embodiments include more than ten D segments. Some preferred embodiments include human D segments.

Typically at least one J_H segment is incorporated in the transgene, although it is preferable to include about six J_H segments, and some preferred embodiments include more than about six J_H segments. Some preferred embodiments include human J_H segments, and further preferred embodiments include six human J_H segments and no nonhuman J_H segments.

S_D segments are donor regions capable of participating in recombination events with the S_A segment of the transgene. For classical isotype switching, S_D and S_A are switch regions such as S_μ , S_{γ_1} , S_{γ_2} , S_{γ_3} , S_{γ_4} , S_α , S_{α_2} , and S_ϵ . Preferably the switch regions are murine or human, more preferably S_D is a human or murine S_μ and S_A is a human or murine S_{γ_1} . For nonclassical isotype switching (δ -associated deletion), S_D and S_A are preferably the 400 basepair direct repeat sequences that flank the human μ gene.

C_1 segments are typically μ or δ genes, preferably a μ gene, and more preferably a human or murine μ gene.

T segments typically include S' flanking sequences that are adjacent to naturally occurring (i.e., germline) switch regions. T segments typically at least about at least 50 nucleotides in length, preferably about at least 200 nucleotides in length, and more preferably at least 500-1000 nucleotides in length. Preferably T segments are 5' flanking sequences that occur immediately upstream of human or murine switch regions in a germline configuration. It is also evident to those of skill in the art that T segments may comprise cis-acting transcriptional regulatory sequences that do not occur naturally in an animal germline (e.g., viral enhancers and promoters such as those found in SV40, adenovirus, and other viruses that infect eukaryotic cells).

C₂ segments are typically a γ_1 , γ_2 , γ_3 , γ_4 , α_1 , α_2 , or ϵ C_H gene, preferably a human C_H gene of these isotypes, and more preferably a human γ_1 or γ_3 gene. Murine γ_{2a} and γ_{2b} may also be used, as may downstream (i.e., switched) isotype genes from various species. Where the heavy chain transgene contains an immunoglobulin heavy chain minilocus, the total length of the transgene will be typically 150 kilo basepairs or less.

In general, the transgene will be other than a native heavy chain Ig locus. Thus, for example, deletion of unnecessary regions or substitutions with corresponding regions from other species will be present.

F. Methods for Determining Functional Isotype Switching in Ig Transgenes

The occurrence of isotype switching in a transgenic nonhuman animal may be identified by any method known to those in the art. Preferred embodiments include the following, employed either singly or in combination:

1. detection of mRNA transcripts that contain a sequence homologous to at least one transgene downstream C_H gene other than δ and an adjacent sequence homologous to a transgene V_H-D_H-J_H rearranged gene; such detection may be by Northern hybridization, S₁ nuclease protection assays, PCR amplification, cDNA cloning, or other methods;

2. detection in the serum of the transgenic animal, or in supernatants of cultures of hybridoma cells made from B-cells of the transgenic animal, of immunoglobulin proteins encoded by downstream C_H genes, where such proteins can also be shown by immunochemical methods to comprise a functional variable region;

3. detection, in DNA from B-cells of the transgenic animal or in genomic DNA from hybridoma cells, of DNA rearrangements consistent with the occurrence of isotype switching in the transgene, such detection may be accomplished by Southern blot hybridization, PCR amplification, genomic cloning, or other method; or

4. identification of other indicia of isotype switching, such as production of sterile transcripts, production of characteristic enzymes involved in switching (e.g., "switch recombinase"), or other manifestations that may be detected, measured, or observed by contemporary techniques.

Because each transgenic line may represent a different site of integration of the transgene, and a potentially different tandem array of transgene inserts, and because each different configuration of transgene and flanking DNA sequences can affect gene expression, it is preferable to identify and use lines of mice that express high levels of human immunoglobulins, particularly of the IgG isotype, and contain the least number of copies of the transgene. Single copy transgenics minimize the potential problem of incomplete allelic expression. Transgenes are typically integrated into host chromosomal DNA, most usually into germline DNA and propagated by subsequent breeding of germline transgenic breeding stock animals. However, other vectors and transgenic methods known in the present art or subsequently developed may be substituted as appropriate and as desired by a practitioner.

Trans-switching to endogenous nonhuman heavy chain constant region genes can occur and produce chimeric heavy chains and antibodies comprising such chimeric human/mouse heavy chains. Such chimeric antibodies may be desired for certain uses described herein or may be undesirable.

G. Functional Disruption of
Endogenous Immunoglobulin Loci

The expression of successfully rearranged immunoglobulin heavy and light transgenes is expected to have a dominant effect by suppressing the rearrangement of the endogenous immunoglobulin genes in the transgenic nonhuman animal. However, another way to generate a nonhuman that is devoid of endogenous antibodies is by mutating the endogenous immunoglobulin loci. Using embryonic stem cell technology and homologous recombination, the endogenous immunoglobulin repertoire can be readily eliminated. The following describes the functional description of the mouse immunoglobulin loci. The vectors and methods disclosed, however, can be readily adapted for use in other non-human animals.

Briefly, this technology involves the inactivation of a gene, by homologous recombination, in a pluripotent cell line that is capable of differentiating into germ cell tissue. A DNA construct that contains an altered, copy of a mouse immunoglobulin gene is introduced into the nuclei of embryonic stem cells. In a portion of the cells, the introduced DNA recombines with the endogenous copy of the mouse gene, replacing it with the altered copy. Cells containing the newly engineered genetic lesion are injected into a host mouse embryo, which is reimplanted into a recipient female. Some of these embryos develop into chimeric mice that possess germ cells entirely derived from the mutant cell line. Therefore, by breeding the chimeric mice it is possible to obtain a new line of mice containing the introduced genetic lesion (reviewed by Capecchi (1989), Science, 244, 1288-1292).

Because the mouse λ locus contributes to only 5% of the immunoglobulins, inactivation of the heavy chain and/or κ -light chain loci is sufficient. There are three ways to disrupt each of these loci, deletion of the J region, deletion of the J-C intron enhancer, and disruption of constant region coding sequences by the introduction of a stop codon. The last option is the most straightforward, in terms of DNA construct design. Elimination of the μ gene disrupts B-cell maturation thereby preventing class switching to any of the

functional heavy chain segments. The strategy for knocking out these loci is outlined below.

To disrupt the mouse μ and κ genes, targeting vectors are used based on the design employed by Jaenisch and co-workers (Zijlstra, et al. (1989), Nature, 342, 435-438) for the successful disruption of the mouse $\beta 2$ -microglobulin gene. The neomycin resistance gene (neo), from the plasmid pMCIneo is inserted into the coding region of the target gene. The pMCIneo insert uses a hybrid viral promoter/enhancer sequence to drive neo expression. This promoter is active in embryonic stem cells. Therefore, neo can be used as a selectable marker for integration of the knock-out construct. The HSV thymidine kinase (tk) gene is added to the end of the construct as a negative selection marker against random insertion events (Zijlstra, et al., supra.).

A preferred strategy for disrupting the heavy chain locus is the elimination of the J region. This region is fairly compact in the mouse, spanning only 1.3 kb. To construct a gene targeting vector, a 15 kb KpnI fragment containing all of the secreted A constant region exons from mouse genomic library is isolated. The 1.3 kb J region is replaced with the 1.1 kb insert from pMCIneo. The HSV tk gene is then added to the 5' end of the KpnI fragment. Correct integration of this construct, via homologous recombination, will result in the replacement of the mouse J_H region with the neo gene. Recombinants are screened by PCR, using a primer based on the neo gene and a primer homologous to mouse sequences 5' of the KpnI site in the D region.

Alternatively, the heavy-chain locus is knocked out by disrupting the coding region of the μ gene. This approach involves the same 15 kb KpnI fragment used in the previous approach. The 1.1 kb insert from pMCIneo is inserted at a unique BamHI site in exon II, and the HSV tk gene added to the 3' KpnI end. Double crossover events on either side of the neo insert, that eliminate the tk gene, are then selected for. These are detected from pools of selected clones by PCR amplification. One of the PCR primers is derived from neo sequences and the other from mouse sequences outside of the

targeting vector. The functional disruption of the mouse immunoglobulin loci is presented in the Examples.

G. Suppressing Expression of Endogenous Immunoglobulin Loci

In addition to functional disruption of endogenous Ig loci, an alternative method for preventing the expression of an endogenous Ig locus is suppression. Suppression of endogenous Ig genes may be accomplished with antisense RNA produced from one or more integrated transgenes, by antisense oligonucleotides, and/or by administration of antisera specific for one or more endogenous Ig chains.

Antisense Polynucleotides

Antisense RNA transgenes can be employed to partially or totally knock-out expression of specific genes (Pepin et al. (1991) Nature 355: 725; Helene., C. and Toulme, J. (1990) Biochimica Biophys. Acta 1049: 99; Stout, J. and Caskey, T. (1990) Somat. Cell Mol. Genet. 16: 369; Munir et al. (1990) Somat. Cell Mol. Genet. 16: 383, each of which is incorporated herein by reference).

"Antisense polynucleotides" are polynucleotides that: (1) are complementary to all or part of a reference sequence, such as a sequence of an endogenous Ig C_H or C_L region, and (2) which specifically hybridize to a complementary target sequence, such as a chromosomal gene locus or a Ig mRNA. Such complementary antisense polynucleotides may include nucleotide substitutions, additions, deletions, or transpositions, so long as specific hybridization to the relevant target sequence is retained as a functional property of the polynucleotide. Complementary antisense polynucleotides include soluble antisense RNA or DNA oligonucleotides which can hybridize specifically to individual mRNA species and prevent transcription and/or RNA processing of the mRNA species and/or translation of the encoded polypeptide (Ching et al., Proc. Natl. Acad. Sci. U.S.A. 86:10006-10010 (1989); Broder et al., Ann. Int. Med.

113:604-618 (1990); Loreau et al., FEBS Letters 274:53-56 (1990); Holcenberg et al., WO91/11535; U.S.S.N. 07/530,165 ("New human CRIPTO gene"); WO91/09865; WO91/04753; WO90/13641; and EP 386563, each of which is incorporated herein by reference). An antisense sequence is a polynucleotide sequence that is complementary to at least one immunoglobulin gene sequence of at least about 15 contiguous nucleotides in length, typically at least 20 to 30 nucleotides in length, and preferably more than about 30 nucleotides in length. However, in some embodiments, antisense sequences may have substitutions, additions, or deletions as compared to the complementary immunoglobulin gene sequence, so long as specific hybridization is retained as a property of the antisense polynucleotide. Generally, an antisense sequence is complementary to an endogenous immunoglobulin gene sequence that encodes, or has the potential to encode after DNA rearrangement, an immunoglobulin chain. In some cases, sense sequences corresponding to an immunoglobulin gene sequence may function to suppress expression, particularly by interfering with transcription.

The antisense polynucleotides therefore inhibit production of the encoded polypeptide(s). In this regard, antisense polynucleotides that inhibit transcription and/or translation of one or more endogenous Ig loci can alter the capacity and/or specificity of a non-human animal to produce immunoglobulin chains encoded by endogenous Ig loci.

Antisense polynucleotides may be produced from a heterologous expression cassette in a transfectant cell or transgenic cell, such as a transgenic pluripotent hematopoietic stem cell used to reconstitute all or part of the hematopoietic stem cell population of an individual, or a transgenic nonhuman animal. Alternatively, the antisense polynucleotides may comprise soluble oligonucleotides that are administered to the external milieu, either in culture medium in vitro or in the circulatory system or interstitial fluid in vivo. Soluble antisense polynucleotides present in the external milieu have been shown to gain access to the cytoplasm and inhibit translation of specific mRNA species. In

some embodiments the antisense polynucleotides comprise methylphosphonate moieties, alternatively phosphorothiolates or O-methylribonucleotides may be used, and chimeric oligonucleotides may also be used (Dagle et al. (1990) Nucleic Acids Res. 18: 4751). For some applications, antisense oligonucleotides may comprise polyamide nucleic acids (Nielsen et al. (1991) Science 254: 1497). For general methods relating to antisense polynucleotides, see Antisense RNA and DNA, (1988), D.A. Melton, Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

Antisense polynucleotides complementary to one or more sequences are employed to inhibit transcription, RNA processing, and/or translation of the cognate mRNA species and thereby effect a reduction in the amount of the respective encoded polypeptide. Such antisense polynucleotides can provide a therapeutic function by inhibiting the formation of one or more endogenous Ig chains in vivo.

Whether as soluble antisense oligonucleotides or as antisense RNA transcribed from an antisense transgene, the antisense polynucleotides of this invention are selected so as to hybridize preferentially to endogenous Ig sequences at physiological conditions in vivo. Most typically, the selected antisense polynucleotides will not appreciably hybridize to heterologous Ig sequences encoded by a heavy or light chain transgene of the invention (i.e., the antisense oligonucleotides will not inhibit transgene Ig expression by more than about 25 to 35 percent).

Antiserum Suppression

Partial or complete suppression of endogenous Ig chain expression can be produced by injecting mice with antisera against one or more endogenous Ig chains (Weiss et al. (1984) Proc. Natl. Acad. Sci. (U.S.A.) 81 211, which is incorporated herein by reference). Antisera are selected so as to react specifically with one or more endogenous (e.g., murine) Ig chains but to have minimal or no cross-reactivity with heterologous Ig chains encoded by an Ig transgene of the invention. Thus, administration of selected antisera

according to a schedule as typified by that of Weiss et al. op.cit. will suppress endogenous Ig chain expression but permits expression of heterologous Ig chain(s) encoded by a transgene of the present invention. Suitable antibody sources for antibody comprise:

(1) monoclonal antibodies, such as a monoclonal antibody that specifically binds to a murine μ , γ , κ , or λ chains but does not react with the human immunoglobulin chain(s) encoded by a human Ig transgene of the invention;

(2) mixtures of such monoclonal antibodies, so that the mixture binds with multiple epitopes on a single species of endogenous Ig chain, with multiple endogenous Ig chains (e.g., murine μ and murine γ , or with multiple epitopes and multiple chains or endogenous immunoglobulins;

(3) polyclonal antiserum or mixtures thereof, typically such antiserum/antisera is monospecific for binding to a single species of endogenous Ig chain (e.g., murine μ , murine γ , murine κ , murine λ) or to multiple species of endogenous Ig chain, and most preferably such antisera possesses negligible binding to human immunoglobulin chains encoded by a transgene of the invention; and/or

(4) a mixture of polyclonal antiserum and monoclonal antibodies binding to a single or multiple species of endogenous Ig chain, and most preferably possessing negligible binding to human immunoglobulin chains encoded by a transgene of the invention. Generally, polyclonal antibodies are preferred, and such substantially monospecific polyclonal antibodies can be advantageously produced from an antiserum raised against human immunoglobulin(s) by pre-adsorption with antibodies derived from the nonhuman animal species (e.g., murine) and/or, for example, by affinity chromatography of the antiserum or purified fraction thereof on an affinity resin containing immobilized human Ig (wherein the bound fraction is enriched for the desired anti-human Ig in the antiserum; the bound fraction is typically eluted with conditions of low pH or a chaotropic salt solution).

Cell separation and/or complement fixation can be employed to provide the enhancement of antibody-directed cell

depletion of lymphocytes expressing endogenous (e.g., murine) immunoglobulin chains. In one embodiment, for example, antibodies are employed for ex vivo depletion of murine Ig-expressing explanted hematopoietic cells and/or B-lineage lymphocytes obtained from a transgenic mouse harboring a human Ig transgene. Thus, hematopoietic cells and/or B-lineage lymphocytes are explanted from a transgenic nonhuman animal harboring a human Ig transgene (preferably harboring both a human heavy chain transgene and a human light chain transgene) and the explanted cells are incubated with an antibody (or antibodies) which (1) binds to an endogenous immunoglobulin (e.g., murine μ and/or κ) and (2) lacks substantial binding to human immunoglobulin chains encoded by the transgene(s). Such antibodies are referred to as "suppression antibodies" for clarity. The explanted cell population is selectively depleted of cells which bind to the suppression antibody(ies); such depletion can be accomplished by various methods, such as (1) physical separation to remove suppression antibody-bound cells from unbound cells (e.g., the suppression antibodies may be bound to a solid support or magnetic bead to immobilize and remove cells binding to the suppression antibody), (2) antibody-dependent cell killing of cells bound by the suppression antibody (e.g., by ADCC, by complement fixation, or by a toxin linked to the suppression antibody), and (3) clonal anergy induced by the suppression antibody, and the like.

Frequently, antibodies used for antibody suppression of endogenous Ig chain production will be capable of fixing complement. It is frequently preferable that such antibodies may be selected so as to react well with a convenient complement source for ex vivo/in vitro depletion, such as rabbit or guinea pig complement. For in vivo depletion, it is generally preferred that the suppressor antibodies possess effector functions in the nonhuman transgenic animal species; thus, a suppression antibody comprising murine effector functions (e.g., ADCC and complement fixation) generally would be preferred for use in transgenic mice.

In one variation, a suppression antibody that specifically binds to a predetermined endogenous immunoglobulin chain is used for ex vivo/in vitro depletion of lymphocytes expressing an endogenous immunoglobulin. A cellular explant (e.g., lymphocyte sample) from a transgenic nonhuman animal harboring a human immunoglobulin transgene is contacted with a suppression antibody and cells specifically binding to the suppression antibody are depleted (e.g., by immobilization, complement fixation, and the like), thus generating a cell subpopulation depleted in cells expressing endogenous (nonhuman) immunoglobulins (e.g., lymphocytes expressing murine Ig). The resultant depleted lymphocyte population (T cells, human Ig-positive B-cells, etc.) can be transferred into a immunocompatible (i.e., MHC-compatible) nonhuman animal of the same species and which is substantially incapable of producing endogenous antibody (e.g., SCID mice, RAG-1 or RAG-2 knockout mice). The reconstituted animal (mouse) can then be immunized with an antigen (or reimmunized with an antigen used to immunize the donor animal from which the explant was obtained) to obtain high-affinity (affinity matured) antibodies and B-cells producing such antibodies. Such B-cells may be used to generate hybridomas by conventional cell fusion and screened. Antibody suppression can be used in combination with other endogenous Ig inactivation/suppression methods (e.g., J_H knockout, C_H knockout, D-region ablation, antisense suppression, compensated frameshift inactivation).

Complete Endogenous Ig Locus Inactivation

In certain embodiments, it is desirable to effect complete inactivation of the endogenous Ig loci so that hybrid immunoglobulin chains comprising a human variable region and a non-human (e.g., murine) constant region cannot be formed (e.g., by trans-switching between the transgene and endogenous Ig sequences). Knockout mice bearing endogenous heavy chain alleles which are functionally disrupted in the J_H region only frequently exhibit trans-switching, typically wherein a rearranged human variable region (VDJ) encoded by a transgene

is expressed as a fusion protein linked to an endogenous murine constant region, although other trans-switched junctions are possible. To overcome this potential problem, it is generally desirable to completely inactivate the endogenous heavy chain locus by any of various methods, including but not limited to the following: (1) functionally disrupting and/or deleting by homologous recombination at least one and preferably all of the endogenous heavy chain constant region genes, (2) mutating at least one and preferably all of the endogenous heavy chain constant region genes to encode a termination codon (or frameshift) to produce a truncated or frameshifted product (if trans-switched), and other methods and strategies apparent to those of skill in the art. Deletion of a substantial portion or all of the heavy chain constant region genes and/or D-region genes may be accomplished by various methods, including sequential deletion by homologous recombination targeting vectors, especially of the "hit-and-run" type and the like. Similarly, functional disruption and/or deletion of at least one endogenous light chain locus (e.g., κ) to ablate endogenous light chain constant region genes is often preferable.

Frequently, it is desirable to employ a frameshifted transgene wherein the heterologous transgene comprises a frameshift in the J segment(s) and a compensating frameshift (i.e., to regenerate the original reading frame) in the initial region (i.e., amino-terminal coding portion) of one or more (preferably all) of the transgene constant region genes. Trans-switching to an endogenous IgH locus constant gene (which does not comprise a compensating frameshift) will result in a truncated or missense product that results in the trans-switched B cell being deleted or non-selected, thus suppressing the trans-switched phenotype.

Antisense suppression and antibody suppression may also be used to effect a substantially complete functional inactivation of endogenous Ig gene product expression (e.g., murine heavy and light chain sequences) and/or trans-switched

antibodies (e.g., human variable/murine constant chimeric antibodies).

Various combinations of the inactivation and suppression strategies may be used to effect essentially total suppression of endogenous (e.g., murine) Ig chain expression.

Trans-Switching

In some variations, it may be desirable to produce a trans-switched immunoglobulin. For example, such trans-switched heavy chains can be chimeric (i.e., a non-murine (human) variable region and a murine constant region). Antibodies comprising such chimeric trans-switched immunoglobulins can be used for a variety of applications where it is desirable to have a non-human (e.g., murine) constant region (e.g., for retention of effector functions in the host, for the presence of murine immunological determinants such as for binding of a secondary antibody which does not bind human constant regions). For one example, a human variable region repertoire may possess advantages as compared to the murine variable region repertoire with respect to certain antigens. Presumably the human V_H , D , J_H , V_L , and J_L genes have been selected for during evolution for their ability to encode immunoglobulins that bind certain evolutionarily important antigens; antigens which provided evolutionary selective pressure for the murine repertoire can be distinct from those antigens which provided evolutionary pressure to shape the human repertoire. Other repertoire advantages may exist, making the human variable region repertoire advantageous when combined with a murine constant region (e.g., a trans-switched murine) isotype. The presence of a murine constant region can afford advantages over a human constant region. For example, a murine γ constant region linked to a human variable region by trans-switching may provide an antibody which possesses murine effector functions (e.g., ADCC, murine complement fixation) so that such a chimeric antibody (preferably monoclonal) which is reactive with a predetermined antigen (e.g., human IL-2 receptor) may be tested in a mouse disease model, such as a mouse model of

graft-versus-host disease wherein the T lymphocytes in the mouse express a functional human IL-2 receptor. Subsequently, the human variable region encoding sequence may be isolated (e.g., by PCR amplification or cDNA cloning from the source (hybridoma clone)) and spliced to a sequence encoding a desired human constant region to encode a human sequence antibody more suitable for human therapeutic uses where immunogenicity is preferably minimized. The polynucleotide(s) having the resultant fully human encoding sequence(s) can be expressed in a host cell (e.g., from an expression vector in a mammalian cell) and purified for pharmaceutical formulation. For some applications, the chimeric antibodies may be used directly without replacing the murine constant region with a human constant region. Other variations and uses of trans-switched chimeric antibodies will be evident to those of skill in the art.

The present invention provides transgenic nonhuman animals containing B lymphocytes which express chimeric antibodies, generally resulting from trans-switching between a human heavy chain transgene and an endogenous murine heavy chain constant region gene. Such chimeric antibodies comprise a human sequence variable region and a murine constant region, generally a murine switched (i.e., non- μ , non- δ) isotype. The transgenic nonhuman animals capable of making chimeric antibodies to a predetermined antigen are usually also competent to make fully human sequence antibodies if both human heavy chain and human light chain transgenes encoding human variable and human constant region genes are integrated. Most typically, the animal is homozygous for a functionally disrupted heavy chain locus and/or light chain locus but retains one or more endogenous heavy chain constant region gene(s) capable of trans-switching (e.g., γ , α , ϵ) and frequently retains a cis-linked enhancer. Such a mouse is immunized with a predetermined antigen, usually in combination with an adjuvant, and an immune response comprising a detectable amount of chimeric antibodies comprising heavy chains composed of human sequence variable regions linked to murine constant region sequences is produced. Typically, the

serum of such an immunized animal can comprise such chimeric antibodies at concentrations of about at least 1 $\mu\text{g/ml}$, often about at least 10 $\mu\text{g/ml}$, frequently at least 30 $\mu\text{g/ml}$, and up to 50 to 100 $\mu\text{g/ml}$ or more. The antiserum containing antibodies comprising chimeric human variable/mouse constant region heavy chains typically also comprises antibodies which comprise human variable/human constant region (complete human sequence) heavy chains. Chimeric trans-switched antibodies usually comprise (1) a chimeric heavy chain composed of a human variable region and a murine constant region (typically a murine gamma) and (2) a human transgene-encoded light chain (typically kappa) or a murine light chain (typically lambda in a kappa knockout background). Such chimeric trans-switched antibodies generally bind to a predetermined antigen (e.g., the immunogen) with an affinity of about at least $1 \times 10^7 \text{ M}^{-1}$, preferably with an affinity of about at least $5 \times 10^7 \text{ M}^{-1}$, more preferably with an affinity of at least $1 \times 10^8 \text{ M}^{-1}$ to $1 \times 10^9 \text{ M}^{-1}$ or more. Frequently, the predetermined antigen is a human protein, such as for example a human cell surface antigen (e.g., CD4, CD8, IL-2 receptor, EGF receptor, PDGF receptor), other human biological macromolecule (e.g., thrombomodulin, protein C, carbohydrate antigen, sialyl Lewis antigen, L-selectin), or nonhuman disease associated macromolecule (e.g., bacterial LPS, virion capsid protein or envelope glycoprotein) and the like.

The invention provides transgenic nonhuman animals comprising a genome comprising: (1) a homozygous functionally disrupted endogenous heavy chain locus comprising at least one murine constant region gene capable of trans-switching (e.g., in cis linkage to a functional switch recombination sequence and typically to a functional enhancer), (2) a human heavy chain transgene capable of rearranging to encode and express a functional human heavy chain variable region and capable of trans-switching (e.g., having a cis-linked RSS); optionally further comprising (3) a human light chain (e.g., kappa) transgene capable of rearranging to encode a functional human light chain variable region and expressing a human sequence light chain; optionally further comprising (4) a homozygous

functionally disrupted endogenous light chain locus (κ , preferably κ and λ); and optionally further comprising (5) a serum comprising an antibody comprising a chimeric heavy chain composed of a human sequence variable region encoded by a human transgene and a murine constant region sequence encoded by an endogenous murine heavy chain constant region gene (e.g., $\gamma 1$, $\gamma 2a$, $\gamma 2b$, $\gamma 3$).

Such transgenic mice may further comprise a serum comprising chimeric antibodies which bind a predetermined human antigen (e.g., CD4, CD8, CEA) with an affinity of about at least $1 \times 10^4 \text{ M}^{-1}$, preferably with an affinity of about at least $5 \times 10^4 \text{ M}^{-1}$, more preferably with an affinity of at least $1 \times 10^7 \text{ M}^{-1}$ to $1 \times 10^9 \text{ M}^{-1}$ or more. Frequently, hybridomas can be made wherein the monoclonal antibodies produced thereby have an affinity of at least $8 \times 10^7 \text{ M}^{-1}$. Chimeric antibodies comprising a heavy chain composed of a murine constant region and a human variable region, often capable of binding to a nonhuman antigen, may also be present in the serum or as an antibody secreted from a hybridoma.

In some variations, it is desirable to generate transgenic mice which have inactivated endogenous mouse heavy chain loci which retain intact heavy chain constant region genes, and which have a human heavy chain transgene capable of trans-switching, and optionally also have a human light chain transgene, optionally with one or more inactivated endogenous mouse light chain loci. Such mice may advantageously produce B cells capable of alternatively expressing antibodies comprising fully human heavy chains and antibodies comprising chimeric (human variable/mouse constant) heavy chains, by trans-switching. The serum of said mice would contain antibodies comprising fully human heavy chains and antibodies comprising chimeric (human variable/mouse constant) heavy chains, preferably in combination with fully human light chains. Hybridomas can be generated from the B cells of said mice.

Generally, such chimeric antibodies can be generated by trans-switching, wherein a human transgene encoding a human variable region (encoded by productive V-D-J rearrangement in

vivo) and a human constant region, typically human μ , undergoes switch recombination with a non-transgene immunoglobulin constant gene switch sequence (RSS) thereby operably linking the transgene-encoded human variable region with a heavy chain constant region which is not encoded by said transgene, typically an endogenous murine immunoglobulin heavy chain constant region or a heterologous (e.g., human) heavy chain constant region encoded on a second transgene. Whereas cis-switching refers to isotype-switching by recombination of RSS elements *within* a transgene, trans-switching involves recombination between a transgene RSS and an RSS element outside the transgene, often on a different chromosome than the chromosome which harbors the transgene.

Trans-switching generally occurs between an RSS of an expressed transgene heavy chain constant region gene and either an RSS of an endogenous murine constant region gene (of a non- μ isotype, typically γ) or an RSS of a human constant region gene contained on a second transgene, often integrated on a separate chromosome.

When trans-switching occurs between an RSS of a first, expressed transgene heavy chain constant region gene (e.g., μ) and an RSS of a human heavy chain constant region gene contained on a second transgene, a non-chimeric antibody having a substantially fully human sequence is produced. For example and not limitation, a polynucleotide encoding a human heavy chain constant region (e.g., $\gamma 1$) and an operably linked RSS (e.g., a $\gamma 1$ RSS) can be introduced (e.g., transfected) into a population of hybridoma cells generated from a transgenic mouse B-cell (or B cell population) expressing an antibody comprising a transgene-encoded human μ chain. The resultant hybridoma cells can be selected for the presence of the introduced polynucleotide and/or for the expression of trans-switched antibody comprising a heavy chain having the variable region (idiotype/antigen reactivity) of the human μ chain and having the constant region encoded by the introduced polynucleotide sequence (e.g., human $\gamma 1$). Trans-switch recombination between the RSS of the transgene-encoded human μ chain and the RSS of the introduced polynucleotide encoding a

downstream isotype (e.g., $\gamma 1$) thereby can generate a trans-switched antibody.

The invention also provides a method for producing such chimeric trans-switched antibodies comprising the step of immunizing with a predetermined antigen a transgenic mouse comprising a genome comprising: (1) a homozygous functionally disrupted endogenous heavy chain locus comprising at least one murine constant region gene capable of trans-switching (e.g., $\gamma 2a$, $\gamma 2b$, $\gamma 1$, $\gamma 3$), (2) a human heavy chain transgene capable of rearranging to encode a functional human heavy chain variable region and expressing a human sequence heavy chain and capable of undergoing isotype switching (and/or trans-switching), and optionally further comprising (3) a human light chain (e.g., kappa) transgene capable of rearranging to encode a functional human light (e.g., kappa) chain variable region and expressing a human sequence light chain, and optionally further comprising (4) a homozygous functionally disrupted endogenous light chain locus (typically κ , preferably both κ and λ), and optionally further comprising (5) a serum comprising an antibody comprising a chimeric heavy chain composed of a human sequence variable region encoded by a human transgene and a murine constant region sequence encoded by an endogenous murine heavy chain constant region gene (e.g., $\gamma 1$, $\gamma 2a$, $\gamma 2b$, $\gamma 3$).

Affinity Tagging: Selecting for Switched Isotypes

Advantageously, trans-switching (and cis-switching) is associated with the process of somatic mutation. Somatic mutation expands the range of antibody affinities encoded by clonal progeny of a B-cell. For example, antibodies produced by hybridoma cells which have undergone switching (trans- or cis-) represent a broader range of antigen-binding affinities than is present in hybridoma cells which have not undergone switching. Thus, a hybridoma cell population (typically clonal) which expresses a first antibody comprising a heavy chain comprising a first human heavy chain variable region in polypeptide linkage to a first human heavy chain constant region (e.g., μ) can be screened for hybridoma cell clonal

variants which express an antibody comprising a heavy chain containing said first human heavy chain variable region in polypeptide linkage to a second heavy chain constant region (e.g., a human γ , α , or ϵ constant region). Such clonal variants can be produced by natural clonal variation producing cis-switching in vitro, by induction of class switching (trans- or cis-) as through the administration of agents that promote isotype switching, such as T-cell-derived lymphokines (e.g., IL-4 and IFN γ), by introduction of a polynucleotide comprising a functional RSS and a heterologous (e.g. human) heavy chain constant region gene to serve as a substrate for trans-switching, or by a combination of the above, and the like. Often, polynucleotides containing a human downstream isotype constant region (e.g., γ 1, γ 3, and the like) with an operably linked RSS will also be introduced into hybridoma cells to promote isotype switching via the trans-switch mechanism.

Class switching and affinity maturation take place within the same population of B cells derived from transgenic animals of the present invention. Therefore, identification of class-switched B cells (or hybridomas derived therefrom) can be used as a screening step for obtaining high affinity monoclonal antibodies. A variety of approaches can be employed to facilitate class switching events such as cis-switching (intratransgene switching), trans-switching, or both. For example, a single continuous human genomic fragment comprising both μ and γ constant region genes with the associated RSS elements and switch regulatory elements (e.g., sterile transcript promoter) can be used as a transgene. However, some portions of the desired single contiguous human genomic fragment can be difficult to clone efficiently, such as due to instability problems when replicated in a cloning host or the like; in particular, the region between δ and γ 3 can prove difficult to clone efficiently, especially as a contiguous fragment comprising the μ gene, γ 3 gene, a V gene, D gene segments, and J gene segments.

Also for example, a discontinuous human transgene (minigene) composed of a human μ gene, human γ 3 gene, a human

V gene(s), human D gene segments, and human J gene segments, with one or more deletions of an intervening (intronic) or otherwise nonessential sequence (e.g., one or more V, D, and/or J segment and/or one or more non- μ constant region gene(s)). Such minigenes have several advantages as compared to isolating a single contiguous segment of genomic DNA spanning all of the essential elements for efficient immunoglobulin expression and switching. For example, such a minigene avoids the necessity of isolating large pieces of DNA which may contain sequences which are difficult to clone (e.g., unstable sequences, poison sequences, and the like). Moreover, miniloci comprising elements necessary for isotype switching (e.g., human γ sterile transcript promoter) for producing cis- or trans-switching, can advantageously undergo somatic mutation and class switching in vivo. As many eukaryotic DNA sequences can prove difficult to clone, omitting non-essential sequences can prove advantageous.

In a variation, hybridoma clones producing antibodies having high binding affinity (e.g., at least 1×10^7 M^{-1} , preferably at least 1×10^8 M^{-1} , more preferably at least 1×10^9 M^{-1} or greater) are obtained by selecting, from a pool of hybridoma cells derived from B cells of transgenic mice harboring a human heavy chain transgene capable of isotype switching (see, supra) and substantially lacking endogenous murine heavy chain loci capable of undergoing productive (in-frame) V-D-J rearrangement, hybridomas which express an antibody comprising a heavy chain comprising a human sequence heavy chain variable region in polypeptide linkage to a human (or mouse) non- μ heavy chain constant region; said antibodies are termed "switched antibodies" as they comprise a "switched heavy chain" which is produced as a consequence of cis-switching and/or trans-switching in vivo or in cell culture. Hybridomas producing switched antibodies generally have undergone the process of somatic mutation, and a pool of said hybridomas will generally have a broader range of antigen binding affinities from which hybridoma clones secreting high affinity antibodies can be selected. Typically, hybridomas secreting a human sequence antibody having substantial binding

affinity (greater than $1 \times 10^7 \text{ M}^{-1}$ to $1 \times 10^8 \text{ M}^{-1}$) for a predetermined antigen and wherein said human sequence antibody comprises human immunoglobulin variable region(s) can be selected by a method comprising a two-step process. One step is to identify and isolate hybridoma cells which secrete immunoglobulins which comprise a switched heavy chain (e.g., by binding hybridoma cells to an immobilized immunoglobulin which specifically binds a switched heavy chain and does not substantially bind to an unswitched isotype, e.g., μ). The other step is to identify hybridoma cells which bind to the predetermined antigen with substantial binding affinity (e.g., by ELISA of hybridoma clone supernatants, FACS analysis using labeled antigen, and the like). Typically, selection of hybridomas which secrete switched antibodies is performed prior to identifying hybridoma cells which bind predetermined antigen. Hybridoma cells which express switched antibodies that have substantial binding affinity for the predetermined antigen are isolated and cultured under suitable growth conditions known in the art, typically as individual selected clones. Optionally, the method comprises the step of culturing said selected clones under conditions suitable for expression of monoclonal antibodies; said monoclonal antibodies are collected and can be administered for therapeutic, prophylactic, and/or diagnostic purposes.

Often, the selected hybridoma clones can serve as a source of DNA or RNA for isolating immunoglobulin sequences which encode immunoglobulins (e.g. a variable region) that bind to (or confer binding to) the predetermined antigen. Subsequently, the human variable region encoding sequence may be isolated (e.g., by PCR amplification or cDNA cloning from the source (hybridoma clone)) and spliced to a sequence encoding a desired human constant region to encode a human sequence antibody more suitable for human therapeutic uses where immunogenicity is preferably minimized. The polynucleotide(s) having the resultant fully human encoding sequence(s) can be expressed in a host cell (e.g., from an expression vector in a mammalian cell) and purified for pharmaceutical formulation.

Xenoenhancers

A heterologous transgene capable of encoding a human immunoglobulin (e.g., a heavy chain) advantageously comprises a cis-linked enhancer which is not derived from the mouse genome, and/or which is not naturally associated in cis with the exons of the heterologous transgene. For example, a human κ transgene (e.g., a κ minilocus) can advantageously comprise a human V_{κ} gene, a human J_{κ} gene, a human C_{κ} gene, and a xenoenhancer, typically said xenoenhancer comprises a human heavy chain intronic enhancer and/or a murine heavy chain intronic enhancer, typically located between a J_{κ} gene and the C_{κ} gene, or located downstream of the C_{κ} gene. For example, the mouse heavy chain J- μ intronic enhancer (Banerji et al. (1983) Cell 33: 729) can be isolated on a 0.9 kb XbaI fragment of the plasmid pKVe2 (see, infra). The human heavy chain J- μ intronic enhancer (Hayday et al. (1984) Nature 307: 334) can be isolated as a 1.4 kb MluI/HindIII fragment (see, infra). Addition of a transcriptionally active xenoenhancer to a transgene, such as a combined xenoenhancer consisting essentially of a human J- μ intronic enhancer linked in cis to a mouse J- μ intronic enhancer, can confer high levels of expression of the transgene, especially where said transgene encodes a light chain, such as human κ . Similarly, a rat 3' enhancer can be advantageously included in a minilocus construct capable of encoding a human heavy chain.

Specific Preferred Embodiments

A preferred embodiment of the invention is an animal containing at least one, typically 2-10, and sometimes 25-50 or more copies of the transgene described in Example 12 (e.g., pHCl or pHc2) bred with an animal containing a single copy of a light chain transgene described in Examples 5, 6, 8, or 14, and the offspring bred with the J_H deleted animal described in Example 10. Animals are bred to homozygosity for each of these three traits. Such animals have the following genotype: a single copy (per haploid set of chromosomes) of a human heavy chain unrearranged mini-locus (described in Example 12),

a single copy (per haploid set of chromosomes) of a rearranged human κ light chain construct (described in Example 14), and a deletion at each endogenous mouse heavy chain locus that removes all of the functional J_H segments (described in Example 10). Such animals are bred with mice that are homozygous for the deletion of the J_H segments (Examples 10) to produce offspring that are homozygous for the J_H deletion and hemizygous for the human heavy and light chain constructs. The resultant animals are injected with antigens and used for production of human monoclonal antibodies against these antigens.

B cells isolated from such an animal are monospecific with regard to the human heavy and light chains because they contain only a single copy of each gene. Furthermore, they will be monospecific with regards to human or mouse heavy chains because both endogenous mouse heavy chain gene copies are nonfunctional by virtue of the deletion spanning the J_H region introduced as described in Example 9 and 12. Furthermore, a substantial fraction of the B cells will be monospecific with regards to the human or mouse light chains because expression of the single copy of the rearranged human κ light chain gene will allelically and isotypically exclude the rearrangement of the endogenous mouse κ and λ chain genes in a significant fraction of B-cells.

The transgenic mouse of the preferred embodiment will exhibit immunoglobulin production with a significant repertoire, ideally substantially similar to that of a native mouse. Thus, for example, in embodiments where the endogenous Ig genes have been inactivated, the total immunoglobulin levels will range from about 0.1 to 10 mg/ml of serum, preferably 0.5 to 5 mg/ml, ideally at least about 1.0 mg/ml. When a transgene capable of effecting a switch to IgG from IgM has been introduced into the transgenic mouse, the adult mouse ratio of serum IgG to IgM is preferably about 10:1. Of course, the IgG to IgM ratio will be much lower in the immature mouse. In general, greater than about 10%, preferably 40 to 80% of the spleen and lymph node B cells express exclusively human IgG protein.

The repertoire will ideally approximate that shown in a non-transgenic mouse, usually at least about 10% as high, preferably 25 to 50% or more. Generally, at least about a thousand different immunoglobulins (ideally IgG), preferably 10^4 to 10^6 or more, will be produced, depending primarily on the number of different V, J and D regions introduced into the mouse genome. These immunoglobulins will typically recognize about one-half or more of highly antigenic proteins, including, but not limited to: pigeon cytochrome C, chicken lysozyme, pokeweed mitogen, bovine serum albumin, keyhole limpet hemocyanin, influenza hemagglutinin, staphylococcus protein A, sperm whale myoglobin, influenza neuraminidase, and lambda repressor protein. Some of the immunoglobulins will exhibit an affinity for preselected antigens of at least about 10^7M^{-1} , preferably 10^8M^{-1} to 10^9M^{-1} or greater.

In some embodiments, it may be preferable to generate mice with predetermined repertoires to limit the selection of V genes represented in the antibody response to a predetermined antigen type. A heavy chain transgene having a predetermined repertoire may comprise, for example, human V_H genes which are preferentially used in antibody responses to the predetermined antigen type in humans. Alternatively, some V_H genes may be excluded from a defined repertoire for various reasons (e.g., have a low likelihood of encoding high affinity V regions for the predetermined antigen; have a low propensity to undergo somatic mutation and affinity sharpening; or are immunogenic to certain humans).

Thus, prior to rearrangement of a transgene containing various heavy or light chain gene segments, such gene segments may be readily identified, e.g. by hybridization or DNA sequencing, as being from a species of organism other than the transgenic animal.

The transgenic mice of the present invention can be immunized with a predetermined antigen, such as a transmembrane proteins, cell surface macromolecule, or other suitable antigen (e.g., TNF, LPS, etc.) for which a human antibody would be desirable. The mice will produce B cells which undergo class-switching via intratransgene switch

recombination (cis-switching) and express immunoglobulins reactive with the predetermined antigen. The immunoglobulins can be human sequence antibodies, wherein the heavy and light chain polypeptides are encoded by human transgene sequences, which may include sequences derived by somatic mutation and V region recombinatorial joints, as well as germline-encoded sequences; these human sequence immunoglobulins can be referred to as being substantially identical to a polypeptide sequence encoded by a human V_L or V_H gene segment and a human J_L or J_H segment, even though other non-germline sequences may be present as a result of somatic mutation and differential V-J and V-D-J recombination joints. With respect to such human sequence antibodies, the variable regions of each chain are typically at least 80 percent encoded by human germline V, J, and, in the case of heavy chains, D, gene segments; frequently at least 85 percent of the variable regions are encoded by human germline sequences present on the transgene; often 90 or 95 percent or more of the variable region sequences are encoded by human germline sequences present on the transgene. However, since non-germline sequences are introduced by somatic mutation and VJ and VDJ joining, the human sequence antibodies will frequently have some variable region sequences (and less frequently constant region sequences) which are not encoded by human V, D, or J gene segments as found in the human transgene(s) in the germline of the mice. Typically, such non-germline sequences (or individual nucleotide positions) will cluster in or near CDRs, or in regions where somatic mutations are known to cluster.

The human sequence antibodies which bind to the predetermined antigen can result from isotype switching, such that human antibodies comprising a human sequence γ chain (such as $\gamma 1$, $\gamma 2a$, $\gamma 2B$, or $\gamma 3$) and a human sequence light chain (such as K) are produced. Such isotype-switched human sequence antibodies often contain one or more somatic mutation(s), typically in the variable region and often in or within about 10 residues of a CDR) as a result of affinity maturation and selection of B cells by antigen, particularly subsequent to secondary (or subsequent) antigen challenge.

These high affinity human sequence antibodies may have binding affinities of at least $1 \times 10^9 \text{ M}^{-1}$, typically at least $5 \times 10^9 \text{ M}^{-1}$, frequently more than $1 \times 10^{10} \text{ M}^{-1}$, and sometimes $5 \times 10^{10} \text{ M}^{-1}$ to 1×10^{11} or greater. Such high affinity human sequence antibodies can be made with high binding affinities for human antigens, such as human CD4 and the like human macromolecules (e.g., such as a human transmembrane or cell surface protein or other cell surface antigen).

The B cells from such mice can be used to generate hybridomas expressing monoclonal high affinity (greater than $2 \times 10^9 \text{ M}^{-1}$) human sequence antibodies against a variety of antigens, including human proteins such as CD4 and the like. These hybridomas can be used to generate a composition comprising an immunoglobulin having an affinity constant (K_a) of at least $2 \times 10^9 \text{ M}^{-1}$ for binding to a predetermined human antigen, wherein said immunoglobulin consists of:

a human sequence light chain composed of (1) a light chain variable region having a polypeptide sequence which is substantially identical to a polypeptide sequence encoded by a human V_L gene segment and a human J_L segment, and (2) a light chain constant region having a polypeptide sequence which is substantially identical to a polypeptide sequence encoded by a human C_L gene segment; and

a human sequence heavy chain composed of a (1) a heavy chain variable region having a polypeptide sequence which is substantially identical to a polypeptide sequence encoded by a human V_H gene segment, optionally a D region, and a human J_H segment, and (2) a constant region having a polypeptide sequence which is substantially identical to a polypeptide sequence encoded by a human C_H gene segment.

Often, the human sequence heavy chain and human sequence light chain are separately encoded by a human heavy chain transgene and a human light chain transgene, respectively, which are integrated into a mouse cell genome. However, both chains may be encoded on a single transgene, or one or both chains may be encoded on multiple transgenes, such as a human heavy chain transgene (e.g., HC2) which derived a V gene segment from a YAC containing a V_H array which is not

integrated at the same locus as the human heavy chain transgene in the mouse germline.

In one embodiment, the composition has an immunoglobulin which comprises a human sequence light chain having a κ constant region and a human sequence heavy chain having a γ constant region.

The mice (and hybridomas derived therefrom) are a source for an immunoglobulin having an affinity constant (K_a) of at least $1 \times 10^{10} \text{ M}^{-1}$ for binding to a predetermined human antigen, wherein said immunoglobulin consists of:

a human sequence light chain composed of (1) a light chain variable region having a polypeptide sequence which is substantially identical to a polypeptide sequence encoded by a human V_L gene segment and a human J_L segment, and (2) a light chain constant region having a polypeptide sequence which is substantially identical to a polypeptide sequence encoded by a human C_L gene segment; and

a human sequence heavy chain composed of (1) a heavy chain variable region having a polypeptide sequence which is substantially identical to a polypeptide sequence encoded by a human V_H gene segment, optionally a D region, and a human J_H segment, and (2) a constant region having a polypeptide sequence which is substantially identical to a polypeptide sequence encoded by a human C_H gene segment.

The invention provides a transgenic mouse comprising: a homozygous pair of functionally disrupted endogenous heavy chain alleles, a homozygous pair of functionally disrupted endogenous light chain alleles, at least one copy of a heterologous immunoglobulin light chain transgene, and at least one copy of a heterologous immunoglobulin heavy chain transgene, and wherein said animal makes an antibody response following immunization with a human antigen wherein the antibody response comprises an immunoglobulin having an affinity constant (K_a) of at least $2 \times 10^9 \text{ M}^{-1}$ for binding to a predetermined human antigen, wherein said immunoglobulin consists of:

a human sequence light chain composed of (1) a light chain variable region having a polypeptide sequence which is

substantially identical to a polypeptide sequence encoded by a human V_L gene segment and a human J_L segment, and (2) a light chain constant region having a polypeptide sequence which is substantially identical to a polypeptide sequence encoded by a human C_L gene segment; and

a human sequence heavy chain composed of a (1) a heavy chain variable region having a polypeptide sequence which is substantially identical to a polypeptide sequence encoded by a human V_H gene segment, optionally a D region, and a human J_H segment, and (2) a constant region having a polypeptide sequence which is substantially identical to a polypeptide sequence encoded by a human C_H gene segment.

Such a transgenic mouse can produce a human sequence immunoglobulin which binds to a human surface or transmembrane protein present on at least one somatic cell type of a human, wherein the immunoglobulin binds said human surface or transmembrane protein with an affinity constant (K_a) of between $1.5 \times 10^9 M^{-1}$ and $1.8 \times 10^{10} M^{-1}$. One example of such a human surface or transmembrane protein is CD4, although others may be used as immunogens as desired.

The development of high affinity human sequence antibodies against predetermined antigens is facilitated by a method for expanding the repertoire of human variable region gene segments in a transgenic mouse having a genome comprising an integrated human immunoglobulin transgene, said method comprising introducing into the genome a V gene transgene comprising V region gene segments which are not present in said integrated human immunoglobulin transgene. Often, the V region transgene is a yeast artificial chromosome comprising a portion of a human V_H or V_L (V_K) gene segment array, as may naturally occur in a human genome or as may be spliced together separately by recombinant methods, which may include out-of-order or omitted V gene segments. Often at least five or more functional V gene segments are contained on the YAC. In this variation, it is possible to make a transgenic mouse produced by the V repertoire expansion method, wherein the mouse expresses an immunoglobulin chain comprising a variable region sequence encoded by a V region gene segment present on

the V region transgene and a C region encoded on the human Ig transgene. By means of the V repertoire expansion method, transgenic mice having at least 5 distinct V genes can be generated; as can mice containing at least about 24 V genes or more. Of course, some V gene segments may be non-functional (e.g., pseudogenes and the like); these segments may be retained or may be selectively deleted by recombinant methods available to the skilled artisan, if desired.

Once the mouse germline has been engineered to contain a functional YAC having an expanded V segment repertoire, substantially not present in the human Ig transgene containing the J and C gene segments, the trait can be propagated and bred into other genetic backgrounds, including backgrounds where the functional YAC having an expanded V segment repertoire is bred into a mouse germline having a different human Ig transgene. Multiple functional YACs having an expanded V segment repertoire may be bred into a germline to work with a human Ig transgene (or multiple human Ig transgenes). Although referred to herein as YAC transgenes, such transgenes when integrated into the genome may substantially lack yeast sequences, such as sequences required for autonomous replication in yeast; such sequences may optionally be removed by genetic engineering (e.g., restriction digestion and pulsed-field gel electrophoresis or other suitable method) after replication in yeast in no longer necessary (i.e., prior to introduction into a mouse ES cell or mouse prozygote).

The invention also provides a method of propagating the trait of human sequence immunoglobulin expression, comprising breeding a transgenic mouse having the human Ig transgene(s), and optionally also having a functional YAC having an expanded V segment repertoire. Both V_H and V_L gene segments may be present on the YAC. The transgenic mouse may be bred into any background desired by the practitioner, including backgrounds harboring other human transgenes, including human Ig transgenes and/or transgenes encoding other human lymphocyte proteins.

The invention also provides a high affinity human sequence immunoglobulin produced by a transgenic mouse having an expanded V region repertoire YAC transgene.

Although the foregoing describes a preferred embodiment of the transgenic animal of the invention, other embodiments are defined by the disclosure herein and more particularly by the transgenes described in the Examples. Four categories of transgenic animal may be defined:

- I. Transgenic animals containing an unrearranged heavy and rearranged light immunoglobulin transgene.
- II. Transgenic animals containing an unrearranged heavy and unrearranged light immunoglobulin transgene
- III. Transgenic animal containing rearranged heavy and an unrearranged light immunoglobulin transgene, and
- IV. Transgenic animals containing rearranged heavy and rearranged light immunoglobulin transgenes.

Of these categories of transgenic animal, the preferred order of preference is as follows $II > I > III > IV$ where the endogenous light chain genes (or at least the κ gene) have been knocked out by homologous recombination (or other method) and $I > II > III > IV$ where the endogenous light chain genes have not been knocked out and must be dominated by allelic exclusion.

As is discussed supra, the invention provides human sequence monoclonal antibodies that are useful in treatment of human diseases. Therapeutic uses of monoclonal antibodies are discussed in, e.g., Larrick and Bourla, *Journal of Biological Response Modifiers*, 5:379-393, which is incorporated herein by reference. Uses of human monoclonal antibodies include treatment of autoimmune diseases, cancer, infectious diseases, transplant rejection, blood disorders such as coagulation disorders, and other diseases.

The antibodies of this invention may be administered to patients by any method known in the medical arts for delivery of proteins. Antibodies are particularly suited for parenteral administration (i.e., subcutaneous, intramuscular or intravenous administration). The pharmaceutical compositions of the present invention are suitable for administration using

alternative drug delivery approaches as well (see, e.g., Langer, *Science*, 249:1527-1533 (1990)).

Pharmaceutical compositions for parenteral administration usually comprise a solution of a monoclonal antibody dissolved in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers can be used, e.g., water, buffered water, 0.4% saline, 0.3% glycine and the like. These solutions are sterile and generally free of particulate matter. These compositions may be sterilized by conventional, well known sterilization techniques. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions such as pH-adjusting and buffering agents, tonicity adjusting agents and the like, for example sodium acetate, sodium chloride, potassium chloride, calcium chloride, sodium lactate, etc. The concentration of antibody in these formulations can vary widely, i.e., from less than about 0.5%, usually at or at least about 0.1% to as much as 1.5% or 2.0% by weight and will be selected primarily based on fluid volumes, viscosities, etc., in accordance with the particular mode of administration selected. Actual methods for preparing parenterally administrable compositions will be known or apparent to those skilled in the art and are described in more detail in, for example, Remington's *Pharmaceutical Sciences*, 17th Ed., Mack Publishing Company, Easton, Pennsylvania (1985), which is incorporated herein by reference.

The compositions containing the present antibodies or a cocktail thereof can be administered for the prophylactic and/or therapeutic treatments. In therapeutic application, compositions are administered to a patient in an amount sufficient to cure or at least partially arrest the infection and its complications. An amount adequate to accomplish this is defined as a "therapeutically effective dose." Amounts effective for this use generally range from about .05 mg/kg body weight to about 5 mg/kg body weight, preferably between about .2 mg/kg body weight to about 1.5 mg/kg body weight.

In some instances it will be desirable to modify the immunoglobulin molecules of the invention to change their

biological activity. For example, the immunoglobulins can be directly or indirectly coupled to other chemotherapeutics agent. A variety of chemotherapeutics can be coupled for targeting. For example, anti-inflammatory agents which may be coupled include immunomodulators, platelet activating factor (PAF) antagonists, cyclooxygenase inhibitors, lipoxxygenase inhibitors, and leukotriene antagonists. Some preferred moieties include cyclosporin A, indomethacin, naproxen, FK-506, mycophenolic acid, and the like. Similarly, anti-oxidants, e.g., superoxide dismutase, are useful in treating reperfusion injury. Likewise, anticancer agents, such as daunomycin, doxorubicin, vinblastine, bleomycin, and the like can be targeted.

The monoclonal antibodies of the invention may also be used to target amphipaths (e.g., liposomes) to sites in a patient. In these preparations, the drug to be delivered is incorporated as part of a liposome in which a human monoclonal antibody is embedded.

The human-sequence monoclonal antibodies of the invention are useful, in part, because they bind specifically to the predetermined antigen against which they are directed. When the predetermined antigen is a human antigen (i.e., a human protein or fragment thereof), it will sometimes be advantageous if the human immunoglobulin of the invention also binds to the *cognate* antigen found in non-human animals, especially animals that are used frequently for drug testing (e.g., preclinical testing of biological activity, pharmacokinetics and safety). These animals include mice, rabbits, rats, dogs, pigs, and, especially, non-human primates such as chimpanzees, apes and monkeys (e.g., Rhesus monkeys and cynomolgus monkeys). The ability to recognize antigens in experimental animals is particularly useful for determining the effect of specific binding on biodistribution of the immunoglobulins. A cognate antigen is an antigen that (i) has a structure (e.g., amino acid sequence) that is substantially similar to the human antigen (i.e., the amino acid sequence of an animal cognate protein will typically be at least about 50% identical to the human protein, usually at least about 70%

identical and often at least about 80% identical or more); (ii) has substantially the same function as the human antigen; and, (iii) often is found in the same cellular compartment as the human antigen. Human and animal cognate antigens typically (but not always) have the same names. Examples of cognate antigens include human tubulin and mouse tubulin, human CD4 and Rhesus CD4, and human IgG and Rat IgG.

An other aspect, the invention provides antigen-binding human mABs comprising at least one polypeptide encoded by an artificial gene. An artificial gene comprises a polypeptide-encoding nucleic acid segment that is synthesized *in vitro* by chemical or enzymatic methods that do not require a cell-derived template nucleic acid strand (e.g., a nucleic acid template obtained from a bacterial cell or an immune or hybridoma cell) and the progeny (through replication) of the artificial gene, i.e., a wholly synthetic nucleic acid.

Although it is routine in genetic engineering to use short synthetic nucleic acids as primers, linkers and the like, it is also possible by chemical and/or enzymatic means to produce wholly synthetic protein-coding nucleic acids that are 30, 50, or more bases in length. The artificial genes of the invention may include both synthetic nucleic acid regions and cell-derived nucleic acid regions. The synthetic nucleic acid region of the artificial gene will generally be at least about 50 bases in length, often at least about 100 bases, typically at least about 200 bases, more often at least about 250 bases and usually over 300 bases or 400 bases in length. Typically the synthetic nucleic acid regions will encode variable gene segments or a portion thereof, e.g., CDR regions, and the constant regions will be encoded by cell-derived nucleic acids. Immunoglobulin polypeptides (i.e., immunoglobulin heavy chains and immunoglobulin light chains) can be conveniently expressed using artificial genes that encode the polypeptides. Usually the artificial genes are operably linked to transcription promoter sequences, e.g., promoter sequences derived from immunoglobulin genes or from viruses (e.g., SV40, CMV, HIV, RSV) or hybrid promoters. The artificial gene may be linked to other sequences as well, e.g.

polyadenylation sequences and introns. One method for expressing an immunoglobulin polypeptide involves insertion of a synthetic nucleic acid encoding one region of an immunoglobulin polypeptide (e.g., a variable region or portion thereof) into a vector that encodes the remaining segments or parts of the immunoglobulin chain (e.g., a μ , γ , γ_2 , γ_3 , γ_4 , δ , ϵ , α_1 or α_2 constant region) and, optionally, promoter (e.g., a CMV (cytomegalovirus) promoter), polyadenylation or other sequences. Such vectors are constructed so that upon introduction into a cell, the cellular transcription and translation of the vector sequences results in an immunoglobulin polypeptide.

Functional human sequence immunoglobulin heavy and light chain genes and polypeptides can be constructed using artificial genes, and used to produce immunoglobulins with a desired specificity such as specific binding to a predetermined antigen. This is accomplished by constructing an artificial gene that encodes an immunoglobulin polypeptide substantially similar to a polypeptide expressed by a cell from, or a hybridoma derived from, a transgenic animal immunized with the predetermined antigen. Thus, the invention provides artificial genes encoding immunoglobulin polypeptides and methods for producing a human-sequence immunoglobulin using an artificial gene(s).

According to this method, a transgenic animal (e.g., a transgenic mouse with a homozygous pair of functionally disrupted endogenous heavy chain alleles, a homozygous pair of functionally disrupted endogenous light chain alleles, at least one copy of a human immunoglobulin light chain transgene, and at least one copy of a human immunoglobulin heavy chain transgene) is immunized with predetermined antigen, e.g., a human protein. Nucleic acid, preferably mRNA, is then collected or isolated from a cell or population of cells in which immunoglobulin gene rearrangement has taken place, and the sequence(s) of nucleic acids encoding the heavy and/or light chains (especially the V segments) of immunoglobulins, or a portion thereof, is determined. This

sequence information is used as a basis for the sequence of the artificial gene.

Sequence determination will generally require isolation of at least a portion of the gene or cDNA of interest, e.g., a portion of a rearranged human transgene or corresponding cDNA encoding an immunoglobulin polypeptide. Usually this requires cloning the DNA or, preferably, mRNA (i.e., cDNA) encoding the human immunoglobulin polypeptide. Cloning is carried out using standard techniques (see, e.g., Sambrook et al. (1989) *Molecular Cloning: A Laboratory Guide*, Vols 1-3, Cold Spring Harbor Press, which is incorporated herein by reference). For example, a cDNA library may be constructed by reverse transcription of polyA⁺ mRNA, preferably membrane-associated mRNA, and the library screened using probes specific for human immunoglobulin polypeptide gene sequences. In a preferred embodiment, however, the polymerase chain reaction (PCR) is used to amplify cDNAs (or portions of full-length cDNAs) encoding an immunoglobulin gene segment of interest (e.g., a light chain variable segment). Because the sequences of the human immunoglobulin polypeptide genes are readily available to those of skill, probes or PCR primers that will specifically hybridize to or amplify a human immunoglobulin gene or segment thereof can be easily designed. See, e.g., Taylor et al., *Nuc. Acids. Res.*, 20:6287 (1992) which is incorporated by reference. Moreover, the sequences of the human transgene of the transgenic mouse will often be known to the practitioner, and primer sequences can be chosen that hybridize to appropriate regions of the transgene. The amplified sequences can be readily cloned into any suitable vector, e.g., expression vectors, minigene vectors, or phage display vectors. It will be appreciated that the particular method of cloning used not critical, so long as it is possible to determine the sequence of some portion of the immunoglobulin polypeptide of interest. As used herein, a nucleic acid that is cloned, amplified, tagged, or otherwise distinguished from background nucleic acids such that the sequence of the nucleic acid of interest can be determined, is considered isolated.

One source for RNA used for cloning and sequencing is a hybridoma produced by obtaining a B cell from the transgenic mouse and fusing the B cell to an immortal cell. An advantage of using hybridomas is that they can be easily screened, and a hybridoma that produces a human monoclonal antibody of interest selected. Alternatively, RNA can be isolated from B cells (or whole spleen) of the immunized animal. When sources other than hybridomas are used, it may be desirable to screen for sequences encoding immunoglobulins or immunoglobulin polypeptides with specific binding characteristics. One method for such screening is the use of phage display technology. Phage display is described in e.g., Dower et al., WO 91/17271, McCafferty et al., WO 92/01047, and Caton and Koprowski, *Proc. Natl. Acad. Sci. USA*, 87:6450-6454 (1990), each of which is incorporated herein by reference. In one embodiment using phage display technology, cDNA from an immunized transgenic mouse (e.g., total spleen cDNA) is isolated, the polymerase chain reaction is used to amplify a cDNA sequences that encode a portion of an immunoglobulin polypeptide, e.g., CDR regions, and the amplified sequences are inserted into a phage vector. cDNAs encoding peptides of interest, e.g., variable region peptides with desired binding characteristics, are identified by standard techniques such as panning.

The sequence of the amplified or cloned nucleic acid is then determined. Typically the sequence encoding an entire variable region of the immunoglobulin polypeptide is determined, however, it will sometimes be adequate to sequence only a portion of a variable region, for example, the CDR-encoding portion. Typically the portion sequenced will be at least 30 bases in length, more often based coding for at least about one-third or at least about one-half of the length of the variable region will be sequenced.

Sequencing can be carried on clones isolated from a cDNA library, or, when PCR is used, after subcloning the amplified sequence or by direct PCR sequencing of the amplified segment. Sequencing is carried out using standard techniques (see, e.g., Sambrook et al. (1989) *Molecular*

Cloning: A Laboratory Guide, Vols 1-3, Cold Spring Harbor Press, and Sanger, F. et al. (1977) Proc. Natl. Acad. Sci. USA 74: 5463-5467, which is incorporated herein by reference). By comparing the sequence of the cloned nucleic acid with published sequences of human immunoglobulin genes and cDNAs, one of skill will readily be able to determine, depending on the region sequenced, (i) the germline segment usage of the hybridoma immunoglobulin polypeptide (including the isotype of the heavy chain) and (ii) the sequence of the heavy and light chain variable regions, including sequences resulting from N-region addition and the process of somatic mutation. One source of immunoglobulin gene sequence information is the National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Md.

In an alternative embodiment, the amino acid sequence of an immunoglobulin of interest may be determined by direct protein sequencing.

An artificial gene can be constructed that has a sequence identical to or substantially similar to, at least a portion of the immunoglobulin-expressing gene (i.e., rearranged transgene). Similarly, the artificial gene can encode an polypeptide that is identical or has substantial similarity to a polypeptide encoded by the sequenced portion of the rearranged transgene. The degeneracy of the genetic code allows the same polypeptide to be encoded by multiple nucleic acid sequences. It is sometimes desirable to change the nucleic acid sequence, for example to introduce restriction sites, change codon usage to reflect a particular expression system, or to remove a glycosylation site. In addition, changes in the hybridoma sequences may be introduced to change the characteristics (e.g., binding characteristics) of the immunoglobulin. For example, changes may be introduced, especially in the CDR regions of the heavy and light chain variable regions, to increase the affinity of the immunoglobulin for the predetermined antigen.

Methods for constructing an synthetic nucleic acids are well known. An entirely chemical synthesis is possible

but in general, a mixed chemical-enzymatic synthesis is carried out in which chemically synthesized oligonucleotides are used in ligation reactions and/or in the polymerase chain reaction to create longer polynucleotides. In a most preferred embodiment, the polymerase chain reaction is carried out using overlapping primers chosen so that the result of the amplification is a DNA with the sequence desired for the artificial gene. The oligonucleotides of the present invention may be synthesized in solid phase or in solution. Generally, solid phase synthesis is preferred. Detailed descriptions of the procedures for solid phase synthesis of oligonucleotides by phosphite-triester, phosphotriester, and H-phosphonate chemistries are widely available. See, for example, Itakura, U.S. Pat. No. 4,401,796; Caruthers et al., U.S. Pat. Nos. 4,458,066 and 4,500,707; Beaucage et al., *Tetrahedron Lett.*, 22:1859-1862; Matteucci et al., *J. Amer. Chem. Soc.*, 103:3185-3191 (1981); Caruthers et al., *Genetic Engineering*, 4:1-17 (1982); Jones, chapter 2, Atkinson et al., chapter 3, and Sproat et al., chapter 4, in Gait, ed. *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Washington, D.C. (1984); Froehler et al., *Tetrahedron Lett.*, 27:469-472 (1986); Froehler et al., *Nucleic Acids Res.*, 14:5399-5407 (1986); Sinha et al., *Tetrahedron Lett.*, 24:5843-5846 (1983); and Sinha et al., *Nucleic Acids Res.*, 12:4539-4557 (1984) which are incorporated herein by reference.

The artificial gene can be introduced into a cell and expressed to produce an immunoglobulin polypeptide. The choice of cell type for expression will depend on many factors (e.g., the level of protein glycosylation desired), but cells capable of secreting human immunoglobulins will be preferred. Especially preferred cells include CHO cells and myeloma-derived cells such as the SP20 and NS0 cell lines. Standard cell culture are well known and are also described in Newman, et al., *Biotechnology*, 10:1455-1460 (1992); Bebbington, et al., *Biotechnology*, 10:169-175 (1992); Cockett, et al., *Biotechnology*, 8:662-667 (1990); Carter, et al., *Biotechnology*, 10:163-167 (1992), each of which is incorporated herein by reference. Methods for introduction of

nucleic acids, e.g., an artificial gene, are well known and include transfection (e.g., by electroporation or liposome-mediated) and transformation. Systems for expression of introduced genes are described generally in Sambrook et al., *supra*.

It is often desirable to express two immunoglobulin polypeptides (i.e., a heavy chain and a light chain) in the same cell so that an immunoglobulin (e.g., an IgG molecule) is produced *in vivo*. Accordingly it will sometimes be desirable to introduce two artificial genes (i.e., one encoding a heavy chain and one encoding a light chain) into a cell. (The two artificial genes can be introduced on a single vector). Alternatively, one artificial gene encoding one immunoglobulin polypeptide can be introduced into a cell that has been genetically engineered to express the other immunoglobulin polypeptide.

It will be apparent that as the cells into which the artificial gene is transfected propagate, the wholly synthetic nucleic acid portion of the artificial gene, will act as a template for replication and transcription. Nonetheless, the progeny genes will have originated from a synthetic nucleic acid (i.e., a polypeptide-encoding nucleic acid molecule that is synthesized *in vitro* by chemical or enzymatic methods that do not require a cell-derived template nucleic acid strand) and as used herein, are also considered artificial genes. Thus, the relationship of the synthetic portion of the artificial gene to the expressed transgene of the hybridoma is one in which there is an *informational* link (i.e., sequence information) but no direct *physical* link.

The invention also provides anti-CD4 monoclonal antibodies useful in therapeutic and diagnostic applications, especially the treatment of human disease. CD4 is a cell surface protein that is expressed primarily on thymocytes and T cells, and which is involved in T-cell function and MHC Class II recognition of antigen. Antibodies directed against CD4 act to reduce the activity of CD4 cells and thus reduce undesirable autoimmune reactions, inflammatory responses and rejection of transplanted organs.

Indeed, administration of anti-CD4 mABs has been shown to prevent (Wofsy, et al., *J. Exp. Med.*, 161:378-391 (1985)) or reverse (Wofsy, et al., *J. Immunol.*, 138:3247-3253 (1987), Waldor, et al., *Science*, 227:415-417 (1985)) autoimmune disease in animal models. Administration of murine or chimeric anti-CD4 mAbs to patients with rheumatoid arthritis has shown evidence of clinical benefit (Knox, et al., *Blood*, 77:20-30 (1991); Goldbery, et al., *J. Autoimmunity*, 4:617-630; Herzog, et al., *Lancet*, ii:1461-1462; Horneff, et al., *Arthritis Rheum.*, 34:129-140; Reiter, et al., *Arthritis Rheum.*, 34:525-536; Wending, et al., *J. Rheum.*, 18:325-327; Van der Lubbe, et al., *Arthritis Rheum.*, 38:1097-1106; Van der Lubbe, et al., *Arthritis Rheum.*, 36:1375-1379; Moreland, et al., *Arthritis Rheum.*, 36:307-318, and Choy, et al., *Arthritis and Rheumatism*, 39(1):52-56 (1996); all of which is incorporated herein by reference). In addition, as noted above, a chimeric anti-CD4 mAB has shown some clinical efficacy in patients with mycosis fungoides (Knox et al. (1991) *Blood* 77:20; which is incorporated herein by reference). Anti-CD4 antibodies are also discussed in Newman, et al., *Biotechnology*, 10:1455-1460 (1992), which is incorporated herein by reference.

EXPERIMENTAL EXAMPLES

METHODS AND MATERIALS

Transgenic mice are derived according to Hogan, et al., "Manipulating the Mouse Embryo: A Laboratory Manual", Cold Spring Harbor Laboratory, which is incorporated herein by reference.

Embryonic stem cells are manipulated according to published procedures (Teratocarcinomas and embryonic stem cells: a practical approach, E.J. Robertson, ed., IRL Press, Washington, D.C., 1987; Zijlstra et al., *Nature* 342:435-438 (1989); and Schwartzberg et al., *Science* 246:799-803 (1989), each of which is incorporated herein by reference).

DNA cloning procedures are carried out according to J. Sambrook, et al. in *Molecular Cloning: A Laboratory Manual*, 2d ed., 1989, Cold Spring Harbor Laboratory Press,

Cold Spring Harbor, N.Y., which is incorporated herein by reference.

Oligonucleotides are synthesized on an Applied Bio Systems oligonucleotide synthesizer according to specifications provided by the manufacturer.

Hybridoma cells and antibodies are manipulated according to "Antibodies: A Laboratory Manual", Ed Harlow and David Lane, Cold Spring Harbor Laboratory (1988), which is incorporated herein by reference.

EXAMPLE 1

Genomic Heavy Chain Human Ig Transgene

This Example describes the cloning and microinjection of a human genomic heavy chain immunoglobulin transgene which is microinjected into a murine zygote.

Nuclei are isolated from fresh human placental tissue as described by Marzluff et al., "Transcription and Translation: A Practical Approach", B.D. Hammes and S.J. Higgins, eds., pp. 89-129, IRL Press, Oxford (1985)). The isolated nuclei (or PBS washed human spermatocytes) are embedded in a low melting point agarose matrix and lysed with EDTA and proteinase K to expose high molecular weight DNA, which is then digested in the agarose with the restriction enzyme NotI as described by M. Finney in Current Protocols in Molecular Biology (F. Ausubel, et al., eds. John Wiley & Sons, Supp. 4, 1988, Section 2.5.1).

The NotI digested DNA is then fractionated by pulsed field gel electrophoresis as described by Anand et al., Nucl. Acids Res. 17:3425-3433 (1989). Fractions enriched for the NotI fragment are assayed by Southern hybridization to detect one or more of the sequences encoded by this fragment. Such sequences include the heavy chain D segments, J segments, μ and $\gamma 1$ constant regions together with representatives of all 6 VH families (although this fragment is identified as 670 kb fragment from HeLa cells by Berman et al. (1988), supra., we have found it to be as 830 kb fragment from human placental an sperm DNA). Those fractions containing this NotI fragment (see Fig. 4) are pooled and cloned into the NotI site of the

vector pYACNN in Yeast cells. Plasmid pYACNN is prepared by digestion of pYAC-4 Neo (Cook et al., Nucleic Acids Res. 16: 11817 (1988)) with EcoRI and ligation in the presence of the oligonucleotide 5' - AAT TGC GGC CGC - 3' (SEQ ID NO: 25).

YAC clones containing the heavy chain NotI fragment are isolated as described by Brownstein et al., Science 244:1348-1351 (1989), and Green et al., Proc. Natl. Acad. Sci. USA 87:1213-1217 (1990), which are incorporated herein by reference. The cloned NotI insert is isolated from high molecular weight yeast DNA by pulse field gel electrophoresis as described by M. Finney, op cit. The DNA is condensed by the addition of 1 mM spermine and microinjected directly into the nucleus of single cell embryos previously described.

EXAMPLE 2

Genomic κ Light Chain Human Ig Transgene Formed by In Vivo Homologous Recombination

A map of the human κ light chain has been described in Lorenz et al., Nucl. Acids Res. 15:9667-9677 (1987), which is incorporated herein by reference.

A 450 kb XhoI to NotI fragment that includes all of C κ , the 3' enhancer, all J segments, and at least five different V segments is isolated and microinjected into the nucleus of single cell embryos as described in Example 1.

EXAMPLE 3

Genomic κ Light Chain Human Ig Transgene Formed by In Vivo Homologous Recombination

A 750 kb MluI to NotI fragment that includes all of the above plus at least 20 more V segments is isolated as described in Example 1 and digested with BssHII to produce a fragment of about 400 kb.

The 450 kb XhoI to NotI fragment plus the approximately 400 kb MluI to BssHII fragment have sequence overlap defined by the BssHII and XhoI restriction sites. Homologous recombination of these two fragments upon microinjection of a mouse zygote results in a transgene

containing at least an additional 15-20 V segments over that found in the 450 kb XhoI/NotI fragment (Example 2).

EXAMPLE 4

Construction of Heavy Chain Mini-Locus

A. Construction of pGP1 and pGP2

pBR322 is digested with EcoRI and StyI and ligated with the following oligonucleotides to generate pGP1 which contains a 147 base pair insert containing the restriction sites shown in Fig. 8. The general overlapping of these oligos is also shown in Fig. 9.

The oligonucleotides are:

- oligo- 5' - CTT GAG CCC GCC TAA TGA GCG GGC TTT
TTT TTG CAT ACT GCG GCC - 3' (SEQ ID NO:26)
- oligo-2 5' - GCA ATG GCC TGG ATC CAT GGC GCG CTA
GCA TCG ATA TCT AGA GCT CGA GCA -3' (SEQ ID NO:27)
- oligo-3 5' - TGC AGA TCT GAA TTC CCG GGT ACC AAG
CTT ACG CGT ACT AGT GCG GCC GCT -3' (SEQ ID NO:28)
- oligo-4 5' - AAT TAG CGG CCG CAC TAG TAC GCG TAA
GCT TGG TAC CCG GGA ATT - 3' (SEQ ID NO:29)
- oligo-5 5' - CAG ATC TGC ATG CTC GAG CTC TAG ATA
TCG ATG CTA GCG CGC CAT GGA TCC - 3' (SEQ ID NO:30)
- oligo-6 5' - AGG CCA TTG CGG CCG CAG TAT GCA AAA
AAA AGC CCG CTC ATT AGG CGG GCT - 3' (SEQ ID NO:31)

This plasmid contains a large polylinker flanked by rare cutting NotI sites for building large inserts that can be isolated from vector sequences for microinjection. The plasmid is based on pBR322 which is relatively low copy compared to the pUC based plasmids (pGP1 retains the pBR322 copy number control region near the origin of replication). Low copy number reduces the potential toxicity of insert sequences. In addition, pGP1 contains a strong transcription terminator sequence derived from trpA (Christie et al., Proc. Natl. Acad. Sci. USA 78:4180 (1981)) inserted between the ampicillin resistance gene and the polylinker. This further

reduces the toxicity associated with certain inserts by preventing readthrough transcription coming from the ampicillin promoters.

Plasmid pGP2 is derived from pGP1 to introduce an additional restriction site (SfiI) in the polylinker. pGP1 is digested with MluI and SpeI to cut the recognition sequences in the polylinker portion of the plasmid.

The following adapter oligonucleotides are ligated to the thus digested pGP1 to form pGP2.

5' CGC GTG GCC GCA ATG GCC A 3' (SEQ ID NO:32)

5' CTA GTG GCC ATT GCG GCC A 3' (SEQ ID NO:33)

pGP2 is identical to pGP1 except that it contains an additional Sfi I site located between the MluI and SpeI sites. This allows inserts to be completely excised with SfiI as well as with NotI.

B. Construction of pRE3 (rat enhancer 3')

An enhancer sequence located downstream of the rat constant region is included in the heavy chain constructs.

The heavy chain region 3' enhancer described by Petterson et al., Nature 344:165-168 (1990), which is incorporated herein by reference) is isolated and cloned. The rat IGH 3' enhancer sequence is PCR amplified by using the following oligonucleotides:

5' CAG GAT CCA GAT ATC AGT ACC TGA AAC AGG GCT TGC
3' (SEQ ID NO:34)

5' GAG CAT GCA CAG GAC CTG GAG CAC ACA CAG CCT TCC
3' (SEQ ID NO:35)

The thus formed double stranded DNA encoding the 3' enhancer is cut with BamHI and SphI and clone into BamHI/SphI cut pGP2 to yield pRE3 (rat enhancer 3').

C. Cloning of Human J- μ Region

A substantial portion of this region is cloned by combining two or more fragments isolated from phage lambda inserts. See Fig. 9.

A 6.3 kb BamHI/HindIII fragment that includes all human J segments (Matsuda et al., EMBO J., 7:1047-1051 (1988); Ravetch et al. Cell, 27:583-591 (1981), which are incorporated herein by reference) is isolated from human genomic DNA library using the oligonucleotide GGA CTG TGT CCC TGT GTG ATG CTT TTG ATG TCT GGG GCC AAG (SEQ ID NO:36).

An adjacent 10 kb HindIII/BamII fragment that contains enhancer, switch and constant region coding exons (Yasui et al., Eur. J. Immunol. 19:1399-1403 (1989)) is similarly isolated using the oligonucleotide: CAC CAA GTT GAC CTG CCT GGT CAC AGA CCT GAC CAC CTA TGA (SEQ ID NO:37).

An adjacent 3' 1.5 kb BamHI fragment is similarly isolated using clone pMUM insert as probe (pMUM is 4 kb EcoRI/HindIII fragment isolated from human genomic DNA library with oligonucleotide:

CCT GTG GAC CAC CGC CTC CAC CTT CAT

CGT CCT CTT CCT CCT (SEQ ID NO:38)

mu membrane exon 1) and cloned into pUC19.

pGP1 is digested with BamHI and BglII followed by treatment with calf intestinal alkaline phosphatase.

Fragments (a) and (b) from Fig. 9 are cloned in the digested pGP1. A clone is then isolated which is oriented such that 5' BamHI site is destroyed by BamHI/Bgl fusion. It is identified as pMU (see Fig. 10). pMU is digested with BamHI and fragment (c) from Fig. 9 is inserted. The orientation is checked with HindIII digest. The resultant plasmid pHIG1 (Fig. 10) contains an 18 kb insert encoding J and C μ segments.

D. Cloning of C_μ Region

pGP1 is digested with BamHI and HindIII is followed by treatment with calf intestinal alkaline phosphatase (Fig. 14). The so treated fragment (b) of Fig. 14 and fragment (c) of Fig. 14 are cloned into the BamHI/HindIII cut pGP1. Proper orientation of fragment (c) is checked by HindIII digestion to form pCON1 containing a 12 kb insert encoding the C_μ region.

Whereas pHIG1 contains J segments, switch and μ sequences in its 18 kb insert with an SfiI 3' site and a SpeI 5' site in a polylinker flanked by NotI sites, will be used for rearranged VDJ segments. pCON1 is identical except that it lacks the J region and contains only a 12 kb insert. The use of pCON1 in the construction of fragment containing rearranged VDJ segments will be described hereinafter.

E. Cloning of γ -1 Constant Region (pREG2)

The cloning of the human γ -1 region is depicted in Fig. 16.

Yamamura et al., Proc. Natl. Acad. Sci. USA 83:2152-2156 (1986) reported the expression of membrane bound human γ -1 from a transgene construct that had been partially deleted on integration. Their results indicate that the 3' BamHI site delineates a sequence that includes the transmembrane rearranged and switched copy of the gamma gene with a V-C intron of less than 5kb. Therefore, in the unrearranged, unswitched gene, the entire switch region is included in a sequence beginning less than 5 kb from the 5' end of the first γ -1 constant exon. Therefore it is included in the 5' 5.3 kb HindIII fragment (Ellison et al., Nucleic Acids Res. 10:4071-4079 (1982), which is incorporated herein by reference). Takahashi et al., Cell 29: 671-679 (1982), which is incorporated herein by reference, also reports that this fragment contains the switch sequence, and this fragment together with the 7.7 kb HindIII to BamHI fragment must include all of the sequences we need for the transgene construct. An intronic sequence is a nucleotide sequence of

at least 15 contiguous nucleotides that occurs in an intron of a specified gene.

Phage clones containing the γ -1 region are identified and isolated using the following oligonucleotide which is specific for the third exon of γ -I (CH3).

5' TGA GCC ACG AAG ACC CTG AGG
TCA AGT TCA ACT GGT ACG TGG 3' (SEQ ID NO:39)

A 7.7 kb HindIII to BglII fragment (fragment (a) in Fig. 11) is cloned into HindIII/BglII cut pRE3 to form pREG1. The upstream 5.3 kb HindIII fragment (fragment (b) in Fig. 11) is cloned into HindIII digested pREG1 to form pREG2. Correct orientation is confirmed by BamHI/SpeI digestion.

F. Combining Cy and Cu

The previously described plasmid pHIG1 contains human J segments and the Cu constant region exons. To provide a transgene containing the Cu constant region gene segments, pHIG1 was digested with SfiI (Fig. 10). The plasmid pREG2 was also digested with SfiI to produce a 13.5 kb insert containing human Cy exons and the rat 3' enhancer sequence. These sequences were combined to produce the plasmid pHIG3' (Fig. 12) containing the human J segments, the human Cu constant region, the human Cy1 constant region and the rat 3' enhancer contained on a 31.5 kb insert.

A second plasmid encoding human Cu and human Cy1 without J segments is constructed by digesting pCON1 with SfiI and combining that with the SfiI fragment containing the human Cy region and the rat 3' enhancer by digesting pREG2 with SfiI. The resultant plasmid, pCON (Fig. 12) contains a 26 kb NotI/SpeI insert containing human Cu, human γ 1 and the rat 3' enhancer sequence.

G. Cloning of D Segment

The strategy for cloning the human D segments is depicted in Fig. 13. Phage clones from the human genomic library containing D segments are identified and isolated

using probes specific for diversity region sequences (Ichihara et al., EMBO J. 7:4141-4150 (1988)). The following oligonucleotides are used:

DXP1: 5' - TGG TAT TAC TAT GGT TCG GGG AGT TAT TAT
AAC CAC AGT GTC - 3' (SEQ ID NO:40)

DXP4: 5' - GCC TGA AAT GGA GCC TCA GGG CAC AGT GGG
CAC GGA CAC TGT - 3' (SEQ ID NO:41)'

DN4: 5' - GCA GGG AGG ACA TGT TTA GGA TCT GAG GCC
GCA CCT GAC ACC - 3' (SEQ ID NO:42)

A 5.2 kb XhoI fragment (fragment (b) in Fig. 13) containing DLR1, DXP1, DXP'1, and DA1 is isolated from a phage clone identified with oligo DXP1.

A 3.2 kb XbaI fragment (fragment (c) in Fig. 13) containing DXP4, DA4 and DK4 is isolated from a phage clone identified with oligo DXP4.

Fragments (b), (c) and (d) from Fig. 13 are combined and cloned into the XbaI/XhoI site of pGP1 to form pHIG2 which contains a 10.6 kb insert.

This cloning is performed sequentially. First, the 5.2 kb fragment (b) in Fig. 13 and the 2.2 kb fragment (d) of Fig. 13 are treated with calf intestinal alkaline phosphatase and cloned into pGP1 digested with XhoI and XbaI. The resultant clones are screened with the 5.2 and 2.2 kb insert. Half of those clones testing positive with the 5.2 and 2.2 kb inserts have the 5.2 kb insert in the proper orientation as determined by BamHI digestion. The 3.2 kb XbaI fragment from Fig. 13 is then cloned into this intermediate plasmid containing fragments (b) and (d) to form pHIG2. This plasmid contains diversity segments cloned into the polylinker with a unique 5' SfiI site and unique 3' SpeI site. The entire polylinker is flanked by NotI sites.

H. Construction of Heavy Chain Minilocus

The following describes the construction of a human heavy chain mini-locus which contain one or more V segments.

An unrearranged V segment corresponding to that identified as the V segment contained in the hybridoma of Newkirk et al., J. Clin. Invest. 81:1511-1518 (1988), which is incorporated herein by reference, is isolated using the following oligonucleotide:

5' - GAT CCT GGT TTA GTT AAA GAG GAT TTT
ATT CAC CCC TGT GTC - 3' (SEQ ID NO:43)

A restriction map of the unrearranged V segment is determined to identify unique restriction sites which provide upon digestion a DNA fragment having a length approximately 2 kb containing the unrearranged V segment together with 5' and 3' flanking sequences. The 5' prime sequences will include promoter and other regulatory sequences whereas the 3' flanking sequence provides recombination sequences necessary for V-DJ joining. This approximately 3.0 kb V segment insert is cloned into the polylinker of pGB2 to form pVH1.

pVH1 is digested with SfiI and the resultant fragment is cloned into the SfiI site of pHIG2 to form a pHIG5'. Since pHIG2 contains D segments only, the resultant pHIG5' plasmid contains a single V segment together with D segments. The size of the insert contained in pHIG5 is 10.6 kb plus the size of the V segment insert.

The insert from pHIG5 is excised by digestion with NotI and SpeI and isolated. pHIG3' which contains J, C_μ and C_γ1 segments is digested with SpeI and NotI and the 3' kb fragment containing such sequences and the rat 3' enhancer sequence is isolated. These two fragments are combined and ligated into NotI digested pGP1 to produce pHIG which contains insert encoding a V segment, nine D segments, six functional J segments, C_μ, C_γ and the rat 3' enhancer. The size of this insert is approximately 43 kb plus the size of the V segment insert.

I. Construction of Heavy Chain Minilocus by Homologous Recombination

As indicated in the previous section, the insert of pHIG is approximately 43 to 45 kb when a single V segment is employed. This insert size is at or near the limit of that which may be readily cloned into plasmid vectors. In order to provide for the use of a greater number of V segments, the following describes in vivo homologous recombination of overlapping DNA fragments which upon homologous recombination within a zygote or ES cell form a transgene containing the rat 3' enhancer sequence, the human C μ , the human C γ 1, human J segments, human D segments and a multiplicity of human V segments.

A 6.3 kb BamHI/HindIII fragment containing human J segments (see fragment (a) in Fig. 9) is cloned into MluI/SpeI digested pHIG5' using the following adapters:

5' GAT CCA AGC AGT 3' (SEQ ID NO:44)

5' CTA GAC TGC TTG 3' (SEQ ID NO:45)

5' CGC GTC GAA CTA 3' (SEQ ID NO:46)

5' AGC TTA GTT CGA 3' (SEQ ID NO:47)

The resultant is plasmid designated pHIG5'O (overlap). The insert contained in this plasmid contains human V, D and J segments. When the single V segment from pVH1 is used, the size of this insert is approximately 17 kb plus 2 kb. This insert is isolated and combined with the insert from pHIG3' which contains the human J, C μ , γ 1 and rat 3' enhancer sequences. Both inserts contain human J segments which provide for approximately 6.3 kb of overlap between the two DNA fragments. When coinjected into the mouse zygote, in vivo homologous recombination occurs generating a transgene equivalent to the insert contained in pHIG.

This approach provides for the addition of a multiplicity of V segments into the transgene formed in vivo. For example, instead of incorporating a single V segment into pHIG5', a multiplicity of V segments contained on (1) isolated

genomic DNA, (2) ligated DNA derived from genomic DNA, or (3) DNA encoding a synthetic V segment repertoire is cloned into pHIG2 at the SfiI site to generate pHIG5' V_N. The J segments fragment (a) of Fig. 9 is then cloned into pHIG5' V_N and the insert isolated. This insert now contains a multiplicity of V segments and J segments which overlap with the J segments contained on the insert isolated from pHIG3'. When cointroduced into the nucleus of a mouse zygote, homologous recombination occurs to generate in vivo the transgene encoding multiple V segments and multiple J segments, multiple D segments, the C_μ region, the C_γ1 region (all from human) and the rat 3' enhancer sequence.

EXAMPLE 5

Construction of Light Chain Minilocus

A. Construction of pEp1

The construction of pEp1 is depicted in Fig. 16. The mouse heavy chain enhancer is isolated on the XbaI to EcoRI 678 bp fragment (Banerji et al., Cell 33:729-740 (1983)) from phage clones using oligo:

5' GAA TGG GAG TGA GGC TCT CTC ATA CCC
TAT TCA GAA CTG ACT 3' (SEQ ID NO:48)

This Ep fragment is cloned into EcoRV/XbaI digested pGP1 by blunt end filling in EcoRI site. The resultant plasmid is designated pEmul.

B. Construction Of κ Light chain Minilocus

The κ construct contains at least one human V_κ segment, all five human J_κ segments, the human J-C_κ enhancer, human κ constant region exon, and, ideally, the human 3' κ enhancer (Meyer et al., EMBO J. 8:1959-1964 (1989)). The κ enhancer in mouse is 9 kb downstream from C_κ. However, it is as yet unidentified in the human. In addition, the construct contains a copy of the mouse heavy chain J-C_μ enhancers.

The minilocus is constructed from four component fragments:

(a) A 16 kb SmaI fragment that contains the human C_κ exon and the 3' human enhancer by analogy with the mouse locus;

(b) A 5' adjacent 5 kb SmaI fragment, which contains all five J segments;

(c) The mouse heavy chain intronic enhancer isolated from pEμ1 (this sequence is included to induce expression of the light chain construct as early as possible in B-cell development. Because the heavy chain genes are transcribed earlier than the light chain genes, this heavy chain enhancer is presumably active at an earlier stage than the intronic κ enhancer); and

(d) A fragment containing one or more V segments.

The preparation of this construct is as follows. Human placental DNA is digested with SmaI and fractionated on agarose gel by electrophoresis. Similarly, human placental DNA is digested with BamHI and fractionated by electrophoresis. The 16 kb fraction is isolated from the SmaI digested gel and the 11 kb region is similarly isolated from the gel containing DNA digested with BamHI.

The 16 kb SmaI fraction is cloned into Lambda FIX II (Stratagene, La Jolla, California) which has been digested with XhoI, treated with klenow fragment DNA polymerase to fill in the XhoI restriction digest product. Ligation of the 16 kb SmaI fraction destroys the SmaI sites and leaves XhoI sites intact.

The 11 kb BamHI fraction is cloned into λ EMBL3 (Stratagene, La Jolla, California) which is digested with BamHI prior to cloning.

Clones from each library were probed with the C_κ specific oligo:

5' GAA CTG TGG CTG CAC CAT CTG TCT
TCA TCT TCC CGC CAT CTG 3' (SEQ ID NO:49)

A 16 kb XhoI insert that was subcloned into the XhoI cut pEμ1 so that C_κ is adjacent to the SmaI site. The resultant plasmid was designated pKap1.

The above C κ specific oligonucleotide is used to probe the λ EMBL3/BamHI library to identify an 11 kb clone. A 5 kb SmaI fragment (fragment (b) in Fig. 20) is subcloned and subsequently inserted into pKap1 digested with SmaI. Those plasmids containing the correct orientation of J segments, C κ and the E μ enhancer are designated pKap2.

One or more V κ segments are thereafter subcloned into the MluI site of pKap2 to yield the plasmid pKapH which encodes the human V κ segments, the human J κ segments, the human C κ segments and the human E μ enhancer. This insert is excised by digesting pKapH with NotI and purified by agarose gel electrophoresis. The thus purified insert is microinjected into the pronucleus of a mouse zygote as previously described.

C. Construction of κ Light Chain Minilocus by In Vivo Homologous Recombination

The 11 kb BamHI fragment is cloned into BamHI digested pGP1 such that the 3' end is toward the SfiI site. The resultant plasmid is designated pKAPint. One or more V κ segments is inserted into the polylinker between the BamHI and SpeI sites in pKAPint to form pKapHV. The insert of pKapHV is excised by digestion with NotI and purified. The insert from pKap2 is excised by digestion with NotI and purified. Each of these fragments contain regions of homology in that the fragment from pKapHV contains a 5 kb sequence of DNA that include the J κ segments which is substantially homologous to the 5 kb SmaI fragment contained in the insert obtained from pKap2. As such, these inserts are capable of homologously recombining when microinjected into a mouse zygote to form a transgene encoding V κ , J κ and C κ .

EXAMPLE 6

Isolation of Genomic Clones Corresponding to Rearranged and Expressed Copies of Immunoglobulin κ Light Chain Genes

This example describes the cloning of immunoglobulin κ light chain genes from cultured cells that express an

immunoglobulin of interest. Such cells may contain multiple alleles of a given immunoglobulin gene. For example, a hybridoma might contain four copies of the κ light chain gene, two copies from the fusion partner cell line and two copies from the original B-cell expressing the immunoglobulin of interest. Of these four copies, only one encodes the immunoglobulin of interest, despite the fact that several of them may be rearranged. The procedure described in this example allows for the selective cloning of the expressed copy of the κ light chain.

A. Double Stranded cDNA

Cells from human hybridoma, or lymphoma, or other cell line that synthesizes either cell surface or secreted or both forms of IgM with a κ light chain are used for the isolation of polyA⁺ RNA. The RNA is then used for the synthesis of oligo dT primed cDNA using the enzyme reverse transcriptase (for general methods see, Goodspeed et al. (1989) Gene 76: 1; Dunn et al. (1989) J. Biol. Chem. 264: 13057). The single stranded cDNA is then isolated and G residues are added to the 3' end using the enzyme polynucleotide terminal transferase. The G-tailed single-stranded cDNA is then purified and used as template for second strand synthesis (catalyzed by the enzyme DNA polymerase) using the following oligonucleotide as a primer:

5' - GAG GTA CAC TGA CAT ACT GGC ATG CCC
CCC CCC CCC - 3' (SEQ ID NO:50)

The double stranded cDNA is isolated and used for determining the nucleotide sequence of the 5' end of the mRNAs encoding the heavy and light chains of the expressed immunoglobulin molecule. Genomic clones of these expressed genes are then isolated. The procedure for cloning the expressed light chain gene is outlined in part B below.

B. Light Chain

The double stranded cDNA described in part A is denatured and used as a template for a third round of DNA synthesis using the following oligonucleotide primer:

5' - GTA CGC CAT ATC AGC TGG ATG AAG TCA TCA GAT
GGC GGG AAG ATG AAG ACA GAT GGT GCA - 3' (SEQ ID NO:51)

This primer contains sequences specific for the constant portion of the κ light chain message (TCA TCA GAT GGC GGG AAG ATG AAG ACA GAT GGT GCA; SEQ ID NO:52) as well as unique sequences that can be used as a primer for the PCR amplification of the newly synthesized DNA strand (GTA CGC CAT ATC AGC TGG ATG AAG; SEQ ID NO:53). The sequence is amplified by PCR using the following two oligonucleotide primers:

5' - GAG GTA CAC TGA CAT ACT GGC ATG -3' (SEQ ID NO:54)
5' - GTA CGC CAT ATC AGC TGG ATG AAG -3' (SEQ ID NO:53)

The PCR amplified sequence is then purified by gel electrophoresis and used as template for dideoxy sequencing reactions using the following oligonucleotide as a primer:

5' - GAG GTA CAC TGA CAT ACT GGC ATG -3' (SEQ ID NO:54)

The first 42 nucleotides of sequence will then be used to synthesize a unique probe for isolating the gene from which immunoglobulin message was transcribed. This synthetic 42 nucleotide segment of DNA will be referred to below as o-kappa.

A Southern blot of DNA, isolated from the Ig expressing cell line and digested individually and in pairwise combinations with several different restriction endonucleases including SmaI, is then probed with the 32-P labelled unique oligonucleotide o-kappa. A unique restriction endonuclease site is identified upstream of the rearranged V segment.

DNA from the Ig expressing cell line is then cut with SmaI and second enzyme (or BamHI or KpnI if there is SmaI

site inside V segment). Any resulting non-blunted ends are treated with the enzyme T4 DNA polymerase to give blunt ended DNA molecules. Then add restriction site encoding linkers (BamHI, EcoRI or XhoI depending on what site does not exist in fragment) and cut with the corresponding linker enzyme to give DNA fragments with BamHI, EcoRI or XhoI ends. The DNA is then size fractionated by agarose gel electrophoresis, and the fraction including the DNA fragment covering the expressed V segment is cloned into lambda EMBL3 or Lambda FIX (Stratagene, La Jolla, California). V segment containing clones are isolated using the unique probe o-kappa. DNA is isolated from positive clones and subcloned into the polylinker of pKap1. The resulting clone is called pRKL.

EXAMPLE 7

Isolation of Genomic Clones Corresponding to Rearranged Expressed Copies of Immunoglobulin Heavy Chain μ Genes

This example describes the cloning of immunoglobulin heavy chain μ genes from cultured cells of expressed and immunoglobulin of interest. The procedure described in this example allows for the selective cloning of the expressed copy of a μ heavy chain gene.

Double-stranded cDNA is prepared and isolated as described herein before. The double-stranded cDNA is denatured and used as a template for a third round of DNA synthesis using the following oligonucleotide primer:

5' - GTA CGC CAT ATC AGC TGG ATG AAG ACA GGA GAC
GAG GGG GAA AAG GGT TGG GGC GGA TGC - 3' (SEQ ID NO:55)

This primer contains sequences specific for the constant portion of the μ heavy chain message (ACA GGA GAC GAG GGG GAA AAG GGT TGG GGC GGA TGC; SEQ ID NO:56) as well as unique sequences that can be used as a primer for the PCR amplification of the newly synthesized DNA strand (GTA CGC CAT ATC AGC TGG ATG AAG; SEQ ID NO:53). The sequence is amplified by PCR using the following two oligonucleotide primers:

5' - GAG GTA CAC TGA CAT ACT GGC ATG - 3' (SEQ ID NO:54)

5' - GTA CTC CAT ATC AGC TGG ATG AAG - 3' (SEQ ID NO:57)

The PCR amplified sequence is then purified by gel electrophoresis and used as template for dideoxy sequencing reactions using the following oligonucleotide as a primer:

5' - GAG GTA CAC TGA CAT ACT GGC ATG - 3' (SEQ ID NO:54)

The first 42 nucleotides of sequence are then used to synthesize a unique probe for isolating the gene from which immunoglobulin message was transcribed. This synthetic 42 nucleotide segment of DNA will be referred to below as o-mu.

A Southern blot of DNA, isolated from the Ig expressing cell line and digested individually and in pairwise combinations with several different restriction endonucleases including MluI (MluI is a rare cutting enzyme that cleaves between the J segment and mu CH1), is then probed with the 32-P labelled unique oligonucleotide o-mu. A unique restriction endonuclease site is identified upstream of the rearranged V segment.

DNA from the Ig expressing cell line is then cut with MluI and second enzyme. MluI or SpeI adapter linkers are then ligated onto the ends and cut to convert the upstream site to MluI or SpeI. The DNA is then size fractionated by agarose gel electrophoresis, and the fraction including the DNA fragment covering the expressed V segment is cloned directly into the plasmid pGPI. V segment containing clones are isolated using the unique probe o-mu, and the insert is subcloned into MluI or MluI/SpeI cut plasmid pCON2. The resulting plasmid is called pRMGH.

EXAMPLE 8

Construction of Human κ Miniloci Transgenes Light Chain Minilocus

A human genomic DNA phage library was screened with kappa light chain specific oligonucleotide probes and isolated clones spanning the J κ -C region. A 5.7 kb ClaI/XhoI fragment

containing $J_{\kappa}1$ together with a 13 kb $XhoI$ fragment containing $J_{\kappa}2-5$ and C_{κ} into pGP1d was cloned and used to create the plasmid pKcor. This plasmid contains $J_{\kappa}1-5$, the kappa intronic enhancer and C_{κ} together with 4.5 kb of 5' and 9 kb of 3' flanking sequences. It also has a unique 5' $XhoI$ site for cloning V_{κ} segments and a unique 3' $SalI$ site for inserting additional cis-acting regulatory sequences.

V kappa genes

A human genomic DNA phage library was screened with V_{κ} light chain specific oligonucleotide probes and isolated clones containing human V_{κ} segments. Functional V segments were identified by DNA sequence analysis. These clones contain TATA boxes, open reading frames encoding leader and variable peptides (including 2 cysteine residues), splice sequences, and recombination heptamer-12 bp spacer-nonamer sequences. Three of the clones were mapped and sequenced. Two of the clones, 65.5 and 65.8 appear to be functional, they contain TATA boxes, open reading frames encoding leader and variable peptides (including 2 cysteine residues), splice sequences, and recombination heptamer-12 bp spacer-nonamer sequences. The third clone, 65.4, appears to encode a $V_{\kappa}I$ pseudogene as it contains a non-canonical recombination heptamer.

One of the functional clones, V_{κ} 65-8, which encodes a $V_{\kappa}III$ family gene, was used to build a light chain minilocus construct.

pKC1

The kappa light chain minilocus transgene pKC1 (Fig. 32) was generated by inserting a 7.5 kb $XhoI/SalI$ fragment containing V_{κ} 65.8 into the 5' $XhoI$ site of pKcor. The transgene insert was isolated by digestion with $NotI$ prior to injection.

The purified insert was microinjected into the pronuclei of fertilized (C57BL/6 x CBA)F2 mouse embryos and transferred the surviving embryos into pseudopregnant females as described by Hogan et al. (in Methods of Manipulating the

Mouse Embryo, 1986, Cold Spring Harbor Laboratory, New York). Mice that developed from injected embryos were analyzed for the presence of transgene sequences by Southern blot analysis of tail DNA. Transgene copy number was estimated by band intensity relative to control standards containing known quantities of cloned DNA. Serum was isolated from these animals and assayed for the presence of transgene encoded human Ig kappa protein by ELISA as described by Harlow and Lane (in Antibodies: A Laboratory Manual, 1988, Cold Spring Harbor Laboratory, New York). Microtiter plate wells were coated with mouse monoclonal antibodies specific for human Ig kappa (clone 6E1, #0173, AMAC, Inc., Westbrook, ME), human IgM (Clone AF6, #0285, AMAC, Inc., Westbrook, ME) and human IgG1 (clone JL512, #0280, AMAC, Inc., Westbrook, ME). Serum samples were serially diluted into the wells and the presence of specific immunoglobulins detected with affinity isolated alkaline phosphatase conjugated goat anti-human Ig (polyvalent) that had been pre-adsorbed to minimize cross-reactivity with mouse immunoglobulins.

Fig. 35 shows the results of an ELISA assay of serum from 8 mice (I.D. #676, 674, 673, 670, 666, 665, 664, and 496). The first seven of these mice developed from embryos that were injected with the pKC1 transgene insert and the eighth mouse is derived from a mouse generated by microinjection of the pHCl transgene (described previously). Two of the seven mice from KC1 injected embryos (I.D.#'s 666 and 664) did not contain the transgene insert as assayed by DAN Southern blot analysis, and five of the mice (I.D.#'s 676, 674, 673, 670, and 665) contained the transgene. All but one of the KC1 transgene positive animals express detectable levels of human Ig kappa protein, and the single non-expressing animal appears to be a genetic mosaic on the basis of DNA Southern blot analysis. The pHCl positive transgenic mouse expresses human IgM and IgG1 but not Ig kappa, demonstrating the specificity of the reagents used in the assay.

pKC2

The kappa light chain minilocus transgene pKC2 was generated by inserting an 8 kb XhoI/SalI fragment containing V_K 65.5 into the 5' XhoI site of pKC1. The resulting transgene insert, which contains two V_K segments, was isolated prior to microinjection by digestion with NotI.

pKVe2

This construct is identical to pKC1 except that it includes 1.2 kb of additional sequence 5' of J_K and is missing 4.5 kb of sequence 3' of V_K 65.8. In addition it contains a 0.9 kb XbaI fragment containing the mouse heavy chain J-μ intronic enhancer (Banerji et al., Cell 33:729-740 (1983)) together with a 1.4 kb MluI/HindIII fragment containing the human heavy chain J-μ intronic enhancer (Hayday et al., Nature 307:334-340 (1984)) inserted downstream. This construct tests the feasibility of initiating early rearrangement of the light chain minilocus to effect allelic and isotypic exclusion. Analogous constructs can be generated with different enhancers, i.e., the mouse or rat 3' kappa or heavy chain enhancer (Meyer and Neuberger, EMBO J. 8:1959-1964 (1989); Petterson et al. Nature 344:165-168 (1990), which are incorporated herein by reference).

Rearranged Light Chain Transgenes

A kappa light chain expression cassette was designed to reconstruct functionally rearranged light chain genes that have been amplified by PCR from human B-cell DNA. The scheme is outlined in Fig. 33. PCR amplified light chain genes are cloned into the vector pK5nx that includes 3.7 kb of 5' flanking sequences isolated from the kappa light chain gene 65.5. The VJ segment fused to the 5' transcriptional sequences are then cloned into the unique XhoI site of the vector pK3ls that includes J_K2-4, the J_K intronic enhancer, C_K, and 9 kb of downstream sequences. The resulting plasmid contains a reconstructed functionally rearranged kappa light chain transgene that can be excised with NotI for microinjection into embryos. The plasmids also contain unique

SalI sites at the 3' end for the insertion of additional cis-acting regulatory sequences.

Two synthetic oligonucleotides (o-130, o-131) were used to amplify rearranged kappa light chain genes from human spleen genomic DNA. Oligonucleotide o-131 (gga ccc aga (g,c)gg aac cat gga a(g,a)(g,a,t,c)) is complementary to the 5' region of V_KIII family light chain genes and overlaps the first ATC of the leader sequence. Oligonucleotide o-130 (gtg caa tca att ctc gag ttt gac tac aga c) is complementary to a sequence approximately 150 bp 3' of J_K1 and includes an XhoI site. These two oligonucleotides amplify a 0.7 kb DNA fragment from human spleen DNA corresponding to rearranged V_KIII genes joined to J_K1 segments. The PCR amplified DNA was digested with NcoI and XhoI and cloned individual PCR products into the plasmid pNN03. The DNA sequence of 5 clones was determined and identified two with functional VJ joints (open reading frames). Additional functionally rearranged light chain clones are collected. The functionally rearranged clones can be individually cloned into light chain expression cassette described above (Fig. 33). Transgenic mice generated with the rearranged light chain constructs can be bred with heavy chain minilocus transgenics to produce a strain of mice that express a spectrum of fully human antibodies in which all of the diversity of the primary repertoire is contributed by the heavy chain. One source of light chain diversity can be from somatic mutation. Because not all light chains will be equivalent with respect to their ability to combine with a variety of different heavy chains, different strains of mice, each containing different light chain constructs can be generated and tested. The advantage of this scheme, as opposed to the use of unrearranged light chain miniloci, is the increased light chain allelic and isotypic exclusion that comes from having the light chain ready to pair with a heavy chain as soon as heavy chain VDJ joining occurs. This combination can result in an increased frequency of B-cells expressing fully human antibodies, and thus it can facilitate the isolation of human Ig expressing hybridomas.

NotI inserts of plasmids pIGM1, pHCl, pIGG1, pKCl, and pKC2 were isolated away from vector sequences by agarose gel electrophoresis. The purified inserts were microinjected into the pronuclei of fertilized (C57BL/6 x CBA)F2 mouse embryos and transferred the surviving embryos into pseudopregnant females as described by Hogan et al. (Hogan et al., Methods of Manipulating the Mouse Embryo, Cold Spring Harbor Laboratory, New York (1986)).

EXAMPLE 9

Inactivation of the Mouse Kappa Light Chain Gene by Homologous Recombination

This example describes the inactivation of the mouse endogenous kappa locus by homologous recombination in embryonic stem (ES) cells followed by introduction of the mutated gene into the mouse germ line by injection of targeted ES cells bearing an inactivated kappa allele into early mouse embryos (blastocysts).

The strategy is to delete J_K and C_K by homologous recombination with a vector containing DNA sequences homologous to the mouse kappa locus in which a 4.5 kb segment of the locus, spanning the J_K gene and C_K segments, is deleted and replaced by the selectable marker neo.

Construction of the kappa targeting vector

The plasmid pGEM7 (KJ1) contains the neomycin resistance gene (neo), used for drug selection of transfected ES cells, under the transcriptional control of the mouse phosphoglycerate kinase (pgk) promoter (XbaI/TaqI fragment; Adra et al. (1987) Gene 60: 65) in the cloning vector pGEM-7Zf(+). The plasmid also includes a heterologous polyadenylation site for the neo gene, derived from the 3' region of the mouse pgk gene (PvuII/HindIII fragment; Boer et al., Biochemical Genetics, 28:299-308 (1990)). This plasmid was used as the starting point for construction of the kappa targeting vector. The first step was to insert sequences homologous to the kappa locus 3' of the neo expression cassette.

Mouse kappa chain sequences (Fig. 20a) were isolated from a genomic phage library derived from liver DNA using oligonucleotide probes specific for the Ck locus:

5' GGC TGA TGC TGC ACC AAC TGT ATC CAT CTT CCC ACC ATC CAG
-3' (SEQ ID NO:58)

and for the Jk5 gene segment:

5' CTC ACG TTC GGT GCT GGG ACC AAG CTG GAG CTG AAA CGT AAG -
3' (SEQ ID NO:59).

An 8 kb BglII/SacI fragment extending 3' of the mouse C_k segment was isolated from a positive phage clone in two pieces, as a 1.2 kb BglII/SacI fragment and a 6.8 kb SacI fragment, and subcloned into BglII/SacI digested pGEM7 (KJ1) to generate the plasmid pNEO-K3' (Fig. 20b).

A 1.2 kb EcoRI/SphI fragment extending 5' of the J_k region was also isolated from a positive phage clone. An SphI/XbaI/BglII/EcoRI adaptor was ligated to the SphI site of this fragment, and the resulting EcoRI fragment was ligated into EcoRI digested pNEO-K3', in the same 5' to 3' orientation as the neo gene and the downstream 3' kappa sequences, to generate pNEO-K5'3' (Fig. 20c).

The Herpes Simplex Virus (HSV) thymidine kinase (TK) gene was then included in the construct in order to allow for enrichment of ES clones bearing homologous recombinants, as described by Mansour et al., Nature 336:348-352 (1988), which is incorporated herein by reference. The HSV TK cassette was obtained from the plasmid pGEM7 (TK), which contains the structural sequences for the HSV TK gene bracketed by the mouse pgk promoter and polyadenylation sequences as described above for pGEM7 (KJ1). The EcoRI site of pGEM7 (TK) was modified to a BamHI site and the TK cassette was then excised as a BamHI/HindIII fragment and subcloned into pGP1b to generate pGP1b-TK. This plasmid was linearized at the XhoI site and the XhoI fragment from pNEO-K5'3', containing the neo gene flanked by genomic sequences from 5' of J_k and 3' of C_k, was inserted into pGP1b-TK to generate the targeting vector

J/C KI (Fig. 20d). The putative structure of the genomic kappa locus following homologous recombination with J/C K1 is shown in Fig. 20e.

Generation and analysis of ES cells with targeted inactivation of a kappa allele

The ES cells used were the AB-1 line grown on mitotically inactive SNL76/7 cell feeder layers (McMahon and Bradley, Cell 62:1073-1085 (1990)) essentially as described (Robertson, E.J. (1987) in Teratocarcinomas and Embryonic Stem Cells: A Practical Approach. E.J. Robertson, ed. (Oxford: IRL Press), p. 71-112). Other suitable ES lines include, but are not limited to, the E14 line (Hooper et al. (1987) Nature 326: 292-295), the D3 line (Doetschman et al. (1985) J. Embryol. Exp. Morph. 87: 27-45), and the CCE line (Robertson et al. (1986) Nature 323: 445-448). The success of generating a mouse line from ES cells bearing a specific targeted mutation depends on the pluripotency of the ES cells (i.e., their ability, once injected into a host blastocyst, to participate in embryogenesis and contribute to the germ cells of the resulting animal).

The pluripotency of any given ES cell line can vary with time in culture and the care with which it has been handled. The only definitive assay for pluripotency is to determine whether the specific population of ES cells to be used for targeting can give rise to chimeras capable of germline transmission of the ES genome. For this reason, prior to gene targeting, a portion of the parental population of AB-1 cells is injected into C57Bl/6J blastocysts to ascertain whether the cells are capable of generating chimeric mice with extensive ES cell contribution and whether the majority of these chimeras can transmit the ES genome to progeny.

The kappa chain inactivation vector J/C K1 was digested with NotI and electroporated into AB-1 cells by the methods described (Hasty et al., Nature, 350:243-246 (1991)). Electroporated cells were plated onto 100 mm dishes at a density of $1-2 \times 10^6$ cells/dish. After 24 hours, G418

(200µg/ml of active component) and FIAU (0.5µM) were added to the medium, and drug-resistant clones were allowed to develop over 10-11 days. Clones were picked, trypsinized, divided into two portions, and further expanded. Half of the cells derived from each clone were then frozen and the other half analyzed for homologous recombination between vector and target sequences.

DNA analysis was carried out by Southern blot hybridization. DNA was isolated from the clones as described (Laird et al., Nucl. Acids Res. 19:4293 (1991)) digested with XbaI and probed with the 800 bp EcoRI/XbaI fragment indicated in Fig. 20e as probe A. This probe detects a 3.7 kb XbaI fragment in the wild type locus, and a diagnostic 1.8 kb band in a locus which has homologously recombined with the targeting vector (see Fig. 20a and e). Of 901 G418 and FIAU resistant clones screened by Southern blot analysis, 7 displayed the 1.8 kb XbaI band indicative of a homologous recombination into one of the kappa genes. These 7 clones were further digested with the enzymes BglII, SacI, and PstI to verify that the vector integrated homologously into one of the kappa genes. When probed with the diagnostic 800 bp EcoRI/XbaI fragment (probe A), BglII, SacI, and PstI digests of wild type DNA produce fragments of 4.1, 5.4, and 7 kb, respectively, whereas the presence of a targeted kappa allele would be indicated by fragments of 2.4, 7.5, and 5.7 kb, respectively (see Fig. 20a and e). All 7 positive clones detected by the XbaI digest showed the expected BglII, SacI, and PstI restriction fragments diagnostic of a homologous recombination at the kappa light chain. In addition, Southern blot analysis of an NsiI digest of the targeted clones using a neo specific probe (probe B, Fig. 20e) generated only the predicted fragment of 4.2 kb, demonstrating that the clones each contained only a single copy of the targeting vector.

Generation of mice bearing the inactivated kappa chain

Five of the targeted ES clones described in the previous section were thawed and injected into C57Bl/6J blastocysts as described (Bradley, A. (1987) in

Teratocarcinomas and Embryonic Stem Cells: A Practical Approach. E.J. Robertson, ed. (Oxford: IRL Press), p. 113-151) and transferred into the uteri of pseudopregnant females to generate chimeric mice resulting from a mixture of cells derived from the input ES cells and the host blastocyst. The extent of ES cell contribution to the chimeras can be visually estimated by the amount of agouti coat coloration, derived from the ES cell line, on the black C57Bl/6J background. Approximately half of the offspring resulting from blastocyst injection of the targeted clones were chimeric (i.e., showed agouti as well as black pigmentation) and of these, the majority showed extensive (70 percent or greater) ES cell contribution to coat pigmentation. The AB1 ES cells are an XY cell line and a majority of these high percentage chimeras were male due to sex conversion of female embryos colonized by male ES cells. Male chimeras derived from 4 of the 5 targeted clones were bred with C57BL/6J females and the offspring monitored for the presence of the dominant agouti coat color indicative of germline transmission of the ES genome. Chimeras from two of these clones consistently generated agouti offspring. Since only one copy of the kappa locus was targeted in the injected ES clones, each agouti pup had a 50 percent chance of inheriting the mutated locus. Screening for the targeted gene was carried out by Southern blot analysis of Bgl II-digested DNA from tail biopsies, using the probe utilized in identifying targeted ES clones (probe A, Fig. 20e). As expected, approximately 50 percent of the agouti offspring showed a hybridizing Bgl II band of 2.4 kb in addition to the wild-type band of 4.1 kb, demonstrating the germline transmission of the targeted kappa locus.

In order to generate mice homozygous for the mutation, heterozygotes were bred together and the kappa genotype of the offspring determined as described above. As expected, three genotypes were derived from the heterozygote matings: wild-type mice bearing two copies of a normal kappa locus, heterozygotes carrying one targeted copy of the kappa gene and one NT kappa gene, and mice homozygous for the kappa mutation. The deletion of kappa sequences from these latter

mice was verified by hybridization of the Southern blots with a probe specific for J_K (probe C, Fig. 20a). Whereas hybridization of the J_K probe was observed to DNA samples from heterozygous and wild-type siblings, no hybridizing signal was present in the homozygotes, attesting to the generation of a novel mouse strain in which both copies of the kappa locus have been inactivated by deletion as a result of targeted mutation.

EXAMPLE 10

Inactivation of the Mouse Heavy Chain Gene by Homologous Recombination

This example describes the inactivation of the endogenous murine immunoglobulin heavy chain locus by homologous recombination in embryonic stem (ES) cells. The strategy is to delete the endogenous heavy chain J segments by homologous recombination with a vector containing heavy chain sequences from which the J_H region has been deleted and replaced by the gene for the selectable marker neo.

Construction of a heavy chain targeting vector

Mouse heavy chain sequences containing the J_H region (Fig. 21a) were isolated from a genomic phage library derived from the D3 ES cell line (Gossler et al., Proc. Natl. Acad. Sci. U.S.A. 83:9065-9069 (1986)) using a J_H4 specific oligonucleotide probe:

5' ACT ATG CTA TGG ACT ACT GGG GTC AAG GAA CCT CAG TCA CCG -3'
(SEQ ID NO:60)

A 3.5 kb genomic SacI/StuI fragment, spanning the J_H region, was isolated from a positive phage clone and subcloned into SacI/SmaI digested pUC18. The resulting plasmid was designated pUC18 J_H. The neomycin resistance gene (neo), used for drug selection of transfected ES cells, was derived from a repaired version of the plasmid pGEM7 (KJ1). A report in the literature (Yenofsky et al. (1990) Proc. Natl. Acad. Sci. (U.S.A.) 87: 3435-3439) documents a point mutation the neo coding sequences of several commonly used expression vectors,

including the construct pMC1neo (Thomas and Cappechi (1987) Cell 51: 503-512) which served as the source of the neo gene used in pGEM7 (KJ1). This mutation reduces the activity of the neo gene product and was repaired by replacing a restriction fragment encompassing the mutation with the corresponding sequence from a wild-type neo clone. The HindIII site in the prepared pGEM7 (KJ1) was converted to a SalI site by addition of a synthetic adaptor, and the neo expression cassette excised by digestion with XbaI/SalI. The ends of the neo fragment were then blunted by treatment with the Klenow form of DNA polI, and the neo fragment was subcloned into the NaeI site of pUC18 J_H, generating the plasmid pUC18 J_H-neo (Fig. 21b).

Further construction of the targeting vector was carried out in a derivative of the plasmid pGP1b. pGP1b was digested with the restriction enzyme NotI and ligated with the following oligonucleotide as an adaptor:

5'- GGC CGC TCG ACG ATA GCC TCG AGG CTA TAA ATC TAG AAG AAT
TCC AGC AAA GCT TTG GC-3' (SEQ ID NO:61)

The resulting plasmid, called pGMT, was used to build the mouse immunoglobulin heavy chain targeting construct.

The Herpes Simplex Virus (HSV) thymidine kinase (TK) gene was included in the construct in order to allow for enrichment of ES clones bearing homologous recombinants, as described by Mansour et al. (Nature 336, 348-352 (1988)). The HSV TK gene was obtained from the plasmid pGEM7 (TK) by digestion with EcoRI and HindIII. The TK DNA fragment was subcloned between the EcoRI and HindIII sites of pGMT, creating the plasmid pGMT-TK (Fig. 21c).

To provide an extensive region of homology to the target sequence, a 5.9 kb genomic XbaI/XhoI fragment, situated 5' of the J_H region, was derived from a positive genomic phage clone by limit digestion of the DNA with XhoI, and partial digestion with XbaI. As noted in Fig. 21a, this XbaI site is not present in genomic DNA, but is rather derived from phage

sequences immediately flanking the cloned genomic heavy chain insert in the positive phage clone. The fragment was subcloned into XbaI/XhoI digested pGMT-TK, to generate the plasmid pGMT-TK-J_H5' (Fig. 21d).

The final step in the construction involved the excision from pUC18 J_H-neo of the 2.8 kb EcoRI fragment which contained the neo gene and flanking genomic sequences 3' of J_H. This fragment was blunted by Klenow polymerase and subcloned into the similarly blunted XhoI site of pGMT-TK-J_H5'. The resulting construct, J_HKO1 (Fig. 21e), contains 6.9 kb of genomic sequences flanking the J_H locus, with a 2.3 kb deletion spanning the J_H region into which has been inserted the neo gene. Fig. 21f shows the structure of an endogenous heavy chain gene after homologous recombination with the targeting construct.

EXAMPLE 11

Generation and analysis of targeted ES cells

AB-1 ES cells (McMahon and Bradley, Cell 62:1073-1085 (1990)) were grown on mitotically inactive SNL76/7 cell feeder layers essentially as described (Robertson, E.J. (1987) Teratocarcinomas and Embryonic Stem Cells: A Practical Approach. E.J. Robertson, ed. (Oxford: IRL Press), pp. 71-112). As described in the previous example, prior to electroporation of ES cells with the targeting construct J_HKO1, the pluripotency of the ES cells was determined by generation of AB-1 derived chimeras which were shown capable of germline transmission of the ES genome.

The heavy chain inactivation vector J_HKO1 was digested with NotI and electroporated into AB-1 cells by the methods described (Hasty et al., Nature 350:243-246 (1991)). Electroporated cells were plated into 100 mm dishes at a density of $1-2 \times 10^6$ cells/dish. After 24 hours, G418 (200mg/ml of active component) and FIAU (0.5mM) were added to the medium, and drug-resistant clones were allowed to develop over 8-10 days. Clones were picked, trypsinized, divided into two portions, and further expanded. Half of the cells derived from each clone were then frozen and the other half analyzed

for homologous recombination between vector and target sequences.

DNA analysis was carried out by Southern blot hybridization. DNA was isolated from the clones as described (Laird et al. (1991) Nucleic Acids Res. 19: 4293), digested with *Stu*I and probed with the 500 bp *Eco*RI/*Stu*I fragment designated as probe A in Fig. 21f. This probe detects a *Stu*I fragment of 4.7 kb in the wild-type locus, whereas a 3 kb band is diagnostic of homologous recombination of endogenous sequences with the targeting vector (see Fig. 21a and f). Of 525 G418 and FIAU doubly-resistant clones screened by Southern blot hybridization, 12 were found to contain the 3 kb fragment diagnostic of recombination with the targeting vector. That these clones represent the expected targeted events at the J_H locus (as shown in Fig. 21f) was confirmed by further digestion with *Hind*III, *Spe*I and *Hpa*I. Hybridization of probe A (see Fig. 21f) to Southern blots of *Hind*III, *Spe*I, and *Hpa*I digested DNA produces bands of 2.3 kb, >10 kb, and >10kb, respectively, for the wild-type locus (see Fig. 21a), whereas bands of 5.3 kb, 3.8 kb, and 1.9 kb, respectively, are expected for the targeted heavy chain locus (see Fig 21f). All 12 positive clones detected by the *Stu*I digest showed the predicted *Hind*III, *Spe*I, and *Hpa*I bands diagnostic of a targeted J_H gene. In addition, Southern blot analysis of a *Stu*I digest of all 12 clones using a neo-specific probe (probe B, Fig. 21f) generated only the predicted fragment of 3 kb, demonstrating that the clones each contained only a single copy of the targeting vector.

Generation of mice carrying the J_H deletion

Three of the targeted ES clones described in the previous section were thawed and injected into C57BL/6J blastocysts as described (Bradley, A. (1987) in Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, E.J. Robertson, ed. (Oxford: IRL Press), p.113-151) and transferred into the uteri of pseudopregnant females. The extent of ES cell contribution to the chimera was visually estimated from the amount of agouti coat coloration, derived

from the ES cell line, on the black C57BL/6J background. Half of the offspring resulting from blastocyst injection of two of the targeted clones were chimeric (i.e., showed agouti as well as black pigmentation); the third targeted clone did not generate any chimeric animals. The majority of the chimeras showed significant (approximately 50 percent or greater) ES cell contribution to coat pigmentation. Since the AB-1 ES cells are an XY cell line, most of the chimeras were male, due to sex conversion of female embryos colonized by male ES cells. Males chimeras were bred with C57BL/6J females and the offspring monitored for the presence of the dominant agouti coat color indicative of germline transmission of the ES genome. Chimeras from both of the clones consistently generated agouti offspring. Since only one copy of the heavy chain locus was targeted in the injected ES clones, each agouti pup had a 50 percent chance of inheriting the mutated locus. Screening for the targeted gene was carried out by Southern blot analysis of *StuI*-digested DNA from tail biopsies, using the probe utilized in identifying targeted ES clones (probe A, Fig. 21f). As expected, approximately 50 percent of the agouti offspring showed a hybridizing *StuI* band of approximately 3 kb in addition to the wild-type band of 4.7 kb, demonstrating germline transmission of the targeted J_H gene segment.

In order to generate mice homozygous for the mutation, heterozygotes were bred together and the heavy chain genotype of the offspring determined as described above. As expected, three genotypes were derived from the heterozygote matings: wild-type mice bearing two copies of the normal J_H locus, heterozygotes carrying one targeted copy of the gene and one normal copy, and mice homozygous for the J_H mutation. The absence of J_H sequences from these latter mice was verified by hybridization of the Southern blots of *StuI*-digested DNA with a probe specific for J_H (probe C, Fig. 21a). Whereas hybridization of the J_H probe to a 4.7 kb fragment in DNA samples from heterozygous and wild-type siblings was observed, no signal was present in samples from the J_H -mutant homozygotes, attesting to the generation of a novel mouse

strain in which both copies of the heavy chain gene have been mutated by deletion of the J_H sequences.

EXAMPLE 12

Heavy Chain Minilocus Transgene

A. Construction of plasmid vectors for cloning large DNA sequences

1. pGP1a

The plasmid pBR322 was digested with EcoRI and StyI and ligated with the following oligonucleotides:

oligo-42 5'- caa gag ccc gcc taa tga gcg ggc ttt ttt ttg cat
act gcg gcc gct -3' (SEQ ID NO:62)

oligo-43 5'- aat tag cgg ccg cag tat gca aaa aaa agc ccg ctc
att agg cgg gct -3' (SEQ ID NO:63)

The resulting plasmid, pGP1a, is designed for cloning very large DNA constructs that can be excised by the rare cutting restriction enzyme NotI. It contains a NotI restriction site downstream (relative to the ampicillin resistance gene, AmpR) of a strong transcription termination signal derived from the trpA gene (Christie et al., Proc. Natl. Acad. Sci. USA 78:4180 (1981)). This termination signal reduces the potential toxicity of coding sequences inserted into the NotI site by eliminating readthrough transcription from the AmpR gene. In addition, this plasmid is low copy relative to the pUC plasmids because it retains the pBR322 copy number control region. The low copy number further reduces the potential toxicity of insert sequences and reduces the selection against large inserts due to DNA replication. The vectors pGP1b, pGP1c, pGP1d, and pGP1f are derived from pGP1a and contain different polylinker cloning sites. The polylinker sequences are given below

pGP1a

NotI
GCGGCCGC

pGP1b

NotI XhoI ClaI BamHI HindIII NotI
 GCggccgcctcgagatcactatcgattaattaaggatccagcagtaagcttgcGGCCGC
 (SEQ ID NO:64)

pGI1c

NotI SmaI XhoI SalI HindIII BamHI SacII NotI
 GCggccgcatcccgggtctcgaggctcgacaagctttcgaggatccgcGGCCGC
 (SEQ ID NO:65)

pGP1d

NotI SalI HindIII ClaI BamHI XhoI NotI
 GCggccgctgtcgacaagcttctcgatggatcctcgagtgcGGCCGC
 (SEQ ID NO:66)

pGP1f

NotI SalI HindIII EcoRI ClaI KpnI BamHI XhoI NotI
 GCggccgctgtcgacaagcttcgaattcagatcgatgtggtacctggatcctcgagtgcGGCCGC
 (SEQ ID NO:67)

Each of these plasmids can be used for the construction of large transgene inserts that are excisable with NotI so that the transgene DNA can be purified away from vector sequences prior to microinjection.

2. pGP1b

pGP1a was digested with NotI and ligated with the following oligonucleotides:

oligo-47 5' - ggc cgc aag ctt act gct gga tcc tta
 att aat cga tag tga tct cga ggc - 3' (SEQ ID NO:68)

oligo-48 5' - ggc cgc ctc gag atc act atc gat taa
 tta agg atc cag cag taa gct tgc - 3' (SEQ ID NO:69)

The resulting plasmid, pGP1b, contains a short polylinker region flanked by NotI sites. This facilitates the construction of large inserts that can be excised by NotI digestion.

3. pGPe

The following oligonucleotides:

oligo-44 5' - ctc cag gat cca gat atc agt acc tga aac agg gct
tgc - 3' (SEQ ID NO:70)

oligo-45 5' - ctc gag cat gca cag gac ctg gag cac aca cag cct
tcc - 3' (SEQ ID NO:71)

were used to amplify the immunoglobulin heavy chain 3' enhancer (S. Petterson, et al., Nature 344:165-168 (1990)) from rat liver DNA by the polymerase chain reaction technique. The amplified product was digested with BamHI and SphI and cloned into BamHI/SphI digested pNNO3 (pNNO3 is a pUC derived plasmid that contains a polylinker with the following restriction sites, listed in order: NotI, BamHI, NcoI, ClaI, EcoRV, XbaI, SacI, XhoI, SphI, PstI, BglII, EcoRI, SmaI, KpnI, HindIII, and NotI). The resulting plasmid, pRE3, was digested with BamHI and HindIII, and the insert containing the rat Ig heavy chain 3' enhancer cloned into BamHI/HindIII digested pGP1b. The resulting plasmid, pGPe (Fig. 22 and Table 1), contains several unique restriction sites into which sequences can be cloned and subsequently excised together with the 3' enhancer by NotI digestion.

TABLE 1

AATTAGCGGCCCGCCTCGAGATCACTATCGATTAATTAAGGATCCAGATATCAGTACCTGAAAC
 AGGGCTGCTCACAACATCTCTCTCTGTCTCTGTCTCTGTGTGTGTCTCTCTGTCTCTGTCTCTCTCT
 GTCTCTCTGTCTCTGTGTGTGTCTCTCTGTCTCTCTCTGTCTCTGTCTCTGTCTCTGTCTCTGTCTCTG
 TCTCTGTCT
 CCGAGTGACTCACTCTGTGCAGGGTTGGCCCTCGGGGCACATGCAAATGGATGTTTGGTTCCATGCAGAAAAACAT
 GTTCTCATTCTCTGAGCCAAAAATAGCATCAATGATTCCCCACCCTGCAGCTGCAGGTTACCCCCACCTGGCC
 AGGTTGACCAGCTTTGGGGATGGGGCTGGGGGTTCATGACCCCTAACGGTGACATTGAATTCAGTGTTTTCCCA
 TTTATCGACACTGCTGGAATCTGACCCTAGGAGGGAATGACAGGAGATAGGCAAGGTCCAAACACCCAGGGAAG
 TGGGAGAGACAGGAAGGCTGTGTGTGTCTCCAGGTCCTGTGCATGCTGCAGATCTGAATTCGGGTACCAAGCTT
 GCGGCCGAGTATGCAAAAAAAGCCCGCTCATTAGGCGGGCTCTTGGCAGAACATATCCATCGCGTCCGCCATC
 TCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCTGGCCACGGGTGCGCATGATCGTGCTCCTGTCTG
 TTGAGGACCCGGCTAGGCTGGCGGGGTGCTTACTGGTTAGCAGAATGAATCACCGATACGCGAGCGAACGTGA
 AGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAACATGAATGGTCTTCGGTTTCGGTGTTTCGTAAAGT
 CTGGAACCGGAAGTCAGCGCCCTGCACCATTAATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCGTG
 GGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCCTGAGTGATTTTTCTCTGGTCCCGCCGCATCCAT
 ACCGCCAGTTGTTTACCCTCACAACGTTCAGTAACCGGGCATGTTTATCATCAGTAACCCGTATCGTCACGATC
 CTCTCTCGTTTTCATCGGTATCATTACCCCCATGAACAGAAATTCCTTACACGGAGGCATCAAGTGACCAAAC
 AGGAAAAACCGCCCTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAACTCAACGAGC
 TGGACGCGGATGAACAGGCAGACATCTGTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCG
 CGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGCTGTAAAGCGG
 ATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGT
 CACGTAGCGATAGCGGAGTGTATACTGGCTTAATATGCGGCATCAGAGCAGATTGTACTGAGAGTGACCATAT
 GCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAAATACCGCATCAGGCGCTCTTCCGCTCCTCGTCACTGA
 CTCGCTGCGCTCGGTCTGTTCCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA
 ATCAGGGGATAACGCAGGAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
 GCTGGCGTTTTTTCATAGGCTCCGCCCCCTGACGAGCATCAGCAAAATCGACGCTCAAGTCAGAGGTGGCGAAA
 CCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTTAGGTATCTCAGTT
 CGGTGTAGGTGCTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTAGCCCGACCGCTGCGCTTATCCG
 GTAACATATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA
 GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAG
 TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
 CCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATC
 CTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTAT
 CAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAA
 CTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCAGGTATCTCAGCGATCTGTCTATTTCTGTTTATCCATAG
 TTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATAC
 CGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTG
 GTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTA
 ATAGTTTGCGCAACGTTGTTGCCATTGCTGCAGGCATCGTGGTGTACGCTCGTCTGTTGGTATGGCTTCATTCA
 GCTCCGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTC
 CTCCGATCGTTGTGAGAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTA
 CTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGATGC
 GGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCA
 TCATTGGAACCGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCCTGTTGAGATCCAGTTTCATGTAACCCA
 CTCGTGCACCAACTGATCTTCAGCATCTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAAACAGGAAGGCAAA
 ATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCTTTTCAATATATTGAA
 GCATTTATCAGGGTTATTGTCTGATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTTC
 CGCGCACATTTCCCGAAAAAGTGCCACCTGACGCTAAGAAACCATTATTATCATGACATTAACCTATAAAAAATA
 GCGGTATCACGAGGCCCTTTCTGCTCTCAAG

Sequence of vector pGPe (SEQ ID NO:72)

B. Construction of IgM expressing minilocus transgene, pIGM1

1. Isolation of J- μ constant region clones and construction of pJM1

A human placental genomic DNA library cloned into the phage vector λ EMBL3/SP6/T7 (Clonetech Laboratories, Inc., Palo Alto, CA) was screened with the human heavy chain J region specific oligonucleotide:

oligo-1 5' -gga ctg tgt ccc tgt gtg atg ctt ttg
atg tct ggg gcc aag -3' (SEQ ID NO:73)

and the phage clone λ 1.3 isolated. A 6 kb HindIII/KpnI fragment from this clone, containing all six J segments as well as D segment DHQ52 and the heavy chain J- μ intronic enhancer, was isolated. The same library was screened with the human μ specific oligonucleotide:

oligo-2 5' - cac caa gtt gac ctg cct ggt cac aga
cct gac cac cta tga - 3' (SEQ ID NO:74)

and the phage clone λ 2.1 isolated. A 10.5 kb HindIII/XhoI fragment, containing the μ switch region and all of the μ constant region exons, was isolated from this clone. These two fragments were ligated together with KpnI/XhoI digested pNNO3 to obtain the plasmid pJM1.

2. pJM2

A 4 kb XhoI fragment was isolated from phage clone λ 2.1 that contains sequences immediately downstream of the sequences in pJM1, including the so called $\Sigma\mu$ element involved in δ -associated deletion of the μ in certain IgD expressing B-cells (Yasui et al., Eur. J. Immunol. 19:1399 (1989), which is incorporated herein by reference). This fragment was treated with the Klenow fragment of DNA polymerase I and ligated to XhoI cut, Klenow treated, pJM1. The resulting plasmid, pJM2 (Fig. 23), had lost the internal XhoI site but retained the 3' XhoI site due to incomplete reaction by the

Klenow enzyme. pJM2 contains the entire human J region, the heavy chain J- μ intronic enhancer, the μ switch region and all of the μ constant region exons, as well as the two 0.4 kb direct repeats, $\sigma\mu$ and $\Sigma\mu$, involved in δ -associated deletion of the μ gene.

3. Isolation of D region clones and construction of pDH1

The following human D region specific oligonucleotide:

oligo-4 5' - tgg tat tac tat ggt tcg ggg agt tat
 tat aac cac agt gtc - 3' (SEQ ID NO:75)

was used to screen the human placenta genomic library for D region clones. Phage clones λ 4.1 and λ 4.3 were isolated. A 5.5 kb XhoI fragment, that includes the D elements D_{K1} , D_{N1} , and D_{M2} (Ichihara et al., EMBO J. 7:4141 (1988)), was isolated from phage clone λ 4.1. An adjacent upstream 5.2 kb XhoI fragment, that includes the D elements D_{LR1} , D_{XP1} , $D_{XP'1}$, and D_{A1} , was isolated from phage clone λ 4.3. Each of these D region XhoI fragments were cloned into the SalI site of the plasmid vector pSP72 (Promega, Madison, WI) so as to destroy the XhoI site linking the two sequences. The upstream fragment was then excised with XhoI and SmaI, and the downstream fragment with EcoRV and XhoI. The resulting isolated fragments were ligated together with SalI digested pSP72 to give the plasmid pDH1. pDH1 contains a 10.6 kb insert that includes at least 7 D segments and can be excised with XhoI (5') and EcoRV (3').

4. pCOR1

The plasmid pJM2 was digested with Asp718 (an isoschizomer of KpnI) and the overhang filled in with the Klenow fragment of DNA polymerase I. The resulting DNA was then digested with ClaI and the insert isolated. This insert was ligated to the XhoI/EcoRV insert of pDH1 and XhoI/ClaI digested pGPe to generate pCOR1 (Fig. 24).

5. pVH251

A 10.3 kb genomic HindIII fragment containing the two human heavy chain variable region segments V_H251 and V_H105 (Humphries et al., Nature 331:446 (1988), which is incorporated herein by reference) was subcloned into pSP72 to give the plasmid pVH251.

6. pIGM1

The plasmid pCOR1 was partially digested with XhoI and the isolated XhoI/SalI insert of pVH251 cloned into the upstream XhoI site to generate the plasmid pIGM1 (Fig. 25). pIGM1 contains 2 functional human variable region segments, at least 8 human D segments all 6 human J_H segments, the human J- μ enhancer, the human $\sigma\mu$ element, the human μ switch region, all of the human μ coding exons, and the human $\Sigma\mu$ element, together with the rat heavy chain 3' enhancer, such that all of these sequence elements can be isolated on a single fragment, away from vector sequences, by digestion with NotI and microinjected into mouse embryo pronuclei to generate transgenic animals.

C. Construction of IgM and IgG expressing minilocus transgene, pHCl

1. Isolation of γ constant region clones

The following oligonucleotide, specific for human IgG ~~Ig-g~~ constant region genes:

oligo-29 5' -cag cag gtg cac acc caa tgc cca tga
gcc cag aca ctg gac- 3' (SEQ ID NO:76)

was used to screen the human genomic library. Phage clones 129.4 and λ 29.5 were isolated. A 4 kb HindIII fragment of phage clone λ 29.4, containing a γ switch region, was used to probe a human placenta genomic DNA library cloned into the phage vector lambda FIXTM II (Stratagene, La Jolla, CA). Phage clone λ Sg1.13 was isolated. To determine the subclass of the different γ clones, dideoxy sequencing reactions were carried out using subclones of each of the three phage clones as templates and the following oligonucleotide as a primer:

oligo-67 5' - tga gcc cag aca ctg gac - 3' (SEQ ID NO:77)

Phage clones λ 29.5 and λ Syl.13 were both determined to be of the γ 1 subclass.

2. pye1

A 7.8 kb HindIII fragment of phage clone λ 29.5, containing the γ 1 coding region was cloned into pUC18. The resulting plasmid, pLT1, was digested with XhoI, Klenow treated, and religated to destroy the internal XhoI site. The resulting clone, pLT1xk, was digested with HindIII and the insert isolated and cloned into pSP72 to generate the plasmid clone pLT1xks. Digestion of pLT1xks at a polylinker XhoI site and a human sequence derived BamHI site generates a 7.6 kb fragment containing the γ 1 constant region coding exons. This 7.6 kb XhoI/BamHI fragment was cloned together with an adjacent downstream 4.5 kb BamHI fragment from phage clone λ 29.5 into XhoI/BamHI digested pGPe to generate the plasmid clone pye1. pye1 contains all of the γ 1 constant region coding exons, together with 5 kb of downstream sequences, linked to the rat heavy chain 3' enhancer.

3. pye2

A 5.3 kb HindIII fragment containing the γ 1 switch region and the first exon of the pre-switch sterile transcript (P. Sideras et al. (1989) International Immunol. 1, 631) was isolated from phage clone λ Syl.13 and cloned into pSP72 with the polylinker XhoI site adjacent to the 5' end of the insert, to generate the plasmid clone pSyls. The XhoI/SalI insert of pSyls was cloned into XhoI digested pye1 to generate the plasmid clone pye2 (Fig. 26). pye2 contains all of the γ 1 constant region coding exons, and the upstream switch region and sterile transcript exons, together with 5 kb of downstream sequences, linked to the rat heavy chain 3' enhancer. This clone contains a unique XhoI site at the 5' end of the insert. The entire insert, together with the XhoI site and the 3' rat

enhancer can be excised from vector sequences by digestion with NotI.

4. pHCl

The plasmid pIGM1 was digested with XhoI and the 43 kb insert isolated and cloned into XhoI digested pge2 to generate the plasmid pHCl (Fig. 25). pHCl contains 2 functional human variable region segments, at least 8 human D segments all 6 human J_H segments, the human J- μ enhancer, the human $\sigma\mu$ element, the human μ switch region, all of the human μ coding exons, the human $\Sigma\mu$ element, and the human $\gamma 1$ constant region, including the associated switch region and sterile transcript associated exons, together with the rat heavy chain 3' enhancer, such that all of these sequence elements can be isolated on a single fragment, away from vector sequences, by digestion with NotI and microinjected into mouse embryo pronuclei to generate transgenic animals.

D. Construction of IgM and IgG expressing minilocus transgene, pHC2

1. Isolation of human heavy chain V region gene VH49.8

The human placental genomic DNA library lambda, FIX™ II, Stratagene, La Jolla, CA) was screened with the following human VH1 family specific oligonucleotide:

oligo-49 5' - gtt aaa gag gat ttt att cac ccc tgt gtc
 ctc tcc aca ggt gtc - 3' (SEQ ID NO:78)

Phage clone λ 49.8 was isolated and a 6.1 kb XbaI fragment containing the variable segment VH49.8 subcloned into pNNO3 (such that the polylinker ClaI site is downstream of VH49.8 and the polylinker XhoI site is upstream) to generate the plasmid pVH49.8. An 800 bp region of this insert was sequenced, and VH49.8 found to have an open reading frame and intact splicing and recombination signals, thus indicating that the gene is functional (Table 2).

141
TABLE 2

TTCCTCAGGC AGGATTTAGG GCTTGGTCTC TCAGCATCCC ACACTTGTAC AGCTGATGTG	60
GCATCTGTGT TTTCTTTCTC ATCCTAGATC AAGCTTTGAG CTGTGAAATA CCCTGCCTCA	120
TGAATATGCA AATAATCTGA GGTCTTCTGA GATAAATATA GATATATTGG TGCCCTGAGA	180
GCATCACATA ACAACCAGAT TCCTCCTCTA AAGAAGCCCC TGGGAGCACA GCTCATCACC	240
ATG GAC TGG ACC TGG AGG TTC CTC TTT GTG GTG GCA GCA GCT ACA G	286
Met Asp Trp Thr Trp Arg Phe Leu Phe Val Val Ala Ala Ala Thr	
1 5 10 15	
GTAAGGGGCT TCCTAGTCCT AAGGCTGAGG AAGGGATCCT GGTTTAGTTA AAGAGGATTT	346
TATTCACCCC TGTGTCCTCT CCACAG GT GTC CAG TCC CAG GTC CAG CTG GTG	398
Gly Val Gln Ser Gln Val Gln Leu Val	
20	
CAG TCT GGG GCT GAG GTG AAG AAG CCT GGG TCC TCG GTG AAG GTC TCC	446
Gln Ser Gly Ala Glu Val Lys Lys Pro Gly Ser Ser Val Lys Val Ser	
25 30 35 40	
TGC AAG GCT TCT GGA GGC ACC TTC AGC AGC TAT GCT ATC AGC TGG GTG	494
Cys Lys Ala Ser Gly Gly Thr Phe Ser Ser Tyr Ala Ile Ser Trp Val	
45 50 55	
CGA CAG GCC CCT GGA CAA GGG CTT GAG TGG ATG GGA AGG ATC ATC CCT	542
Arg Gln Ala Pro Gly Gln Gly Leu Glu Trp Met Gly Arg Ile Ile Pro	
60 65 70	
ATC CTT GGT ATA GCA AAC TAC GCA CAG AAG TTC CAG GGC AGA GTC ACG	590
Ile Leu Gly Ile Ala Asn Tyr Ala Gln Lys Phe Gln Gly Arg Val Thr	
75 80 85	
ATT ACC GCG GAC AAA TCC ACG AGC ACA GCC TAC ATG GAG CTG AGC AGC	638
Ile Thr Ala Asp Lys Ser Thr Ser Thr Ala Tyr Met Glu Leu Ser Ser	
90 95 100	
CTG AGA TCT GAG GAC ACG GCC GTG TAT TAC TGT GCG AGA GACACAGTGT	687
Leu Arg Ser Glu Asp Thr Ala Val Tyr Tyr Cys Ala Arg	
105 110 115	
GAAAACCCAC ATCCTGAGAG TGTCAGAAAC CCTGAGGGAG AAGGCAGCTG TGCCGGGCTG	747
AGGAGATGAC AGGGTTTATT AGGTTTAAGG CTGTTTACAA AATGGGTAT ATATTTGAGA	807
AAAAA	812

Sequence of human V_HI family gene V_H 49.8 (SEQ ID NOS:79 and 80)

2. pV2

A 4 kb XbaI genomic fragment containing the human V_HIV family gene V_H4-21 (Sanz et al., EMBO J., 8:3741 (1989)), subcloned into the plasmid pUC12, was excised with SmaI and HindIII, and treated with the Klenow fragment of polymerase I. The blunt ended fragment was then cloned into ClaI digested, Klenow treated, pVH49.8. The resulting plasmid, pV2, contains the human heavy chain gene VH49.8 linked upstream of VH4-21 in the same orientation, with a unique SalI site at the 3' end of the insert and a unique XhoI site at the 5' end.

3. pSy1-5'

A 0.7 kb XbaI/HindIII fragment (representing sequences immediately upstream of, and adjacent to, the 5.3 kb γ 1 switch region containing fragment in the plasmid p γ e2) together with the neighboring upstream 3.1 kb XbaI fragment were isolated from the phage clone λ Sg1.13 and cloned into HindIII/XbaI digested pUC18 vector. The resulting plasmid, pSy1-5', contains a 3.8 kb insert representing sequences upstream of the initiation site of the sterile transcript found in B-cells prior to switching to the γ 1 isotype (P. Sideras et al., International Immunol. 1:631 (1989)). Because the transcript is implicated in the initiation of isotype switching, and upstream cis-acting sequences are often important for transcription regulation, these sequences are included in transgene constructs to promote correct expression of the sterile transcript and the associated switch recombination.

4. pVGE1

The pSy1-5' insert was excised with SmaI and HindIII, treated with Klenow enzyme, and ligated with the following oligonucleotide linker:

5' - ccg gtc gac cgg - 3' (SEQ ID NO:81)

The ligation product was digested with SalI and ligated to SalI digested pV2. The resulting plasmid, pVP, contains 3.8

kb of $\gamma 1$ switch 5' flanking sequences linked downstream of the two human variable gene segments VH49.8 and VH4-21 (see Table 2). The pVP insert is isolated by partial digestion with SalI and complete digestion with XhoI, followed by purification of the 15 kb fragment on an agarose gel. The insert is then cloned into the XhoI site of p γ e2 to generate the plasmid clone pVGE1 (Fig. 27). pVGE1 contains two human heavy chain variable gene segments upstream of the human $\gamma 1$ constant gene and associated switch region. A unique SalI site between the variable and constant regions can be used to clone in D, J, and μ gene segments. The rat heavy chain 3' enhancer is linked to the 3' end of the $\gamma 1$ gene and the entire insert is flanked by NotI sites.

5. pHC2

The plasmid clone pVGE1 is digested with SalI and the XhoI insert of pIGM1 is cloned into it. The resulting clone, pHC2 (Fig. 25), contains 4 functional human variable region segments, at least 8 human D segments all 6 human J_H segments, the human J-m enhancer, the human μ element, the human μ switch region, all of the human μ coding exons, the human $\Sigma\mu$ element, and the human $\gamma 1$ constant region, including the associated switch region and sterile transcript associated exons, together with 4 kb flanking sequences upstream of the sterile transcript initiation site. These human sequences are linked to the rat heavy chain 3' enhancer, such that all of the sequence elements can be isolated on a single fragment, away from vector sequences, by digestion with NotI and microinjected into mouse embryo pronuclei to generate transgenic animals. A unique XhoI site at the 5' end of the insert can be used to clone in additional human variable gene segments to further expand the recombinational diversity of this heavy chain minilocus.

E. Transgenic mice

The NotI inserts of plasmids pIGM1 and pHC1 were isolated from vector sequences by agarose gel electrophoresis. The purified inserts were microinjected into the pronuclei of

fertilized (C57BL/6 x CBA)F2 mouse embryos and transferred the surviving embryos into pseudopregnant females as described by Hogan et al. (B. Hogan, F. Costantini, and E. Lacy, *Methods of Manipulating the Mouse Embryo*, 1986, Cold Spring Harbor Laboratory, New York). Mice that developed from injected embryos were analyzed for the presence of transgene sequences by Southern blot analysis of tail DNA. Transgene copy number was estimated by band intensity relative to control standards containing known quantities of cloned DNA. At 3 to 8 weeks of age, serum was isolated from these animals and assayed for the presence of transgene encoded human IgM and IgG1 by ELISA as described by Harlow and Lane (E. Harlow and D. Lane.

Antibodies: A Laboratory Manual, 1988, Cold Spring Harbor Laboratory, New York). Microtiter plate wells were coated with mouse monoclonal antibodies specific for human IgM (clone AF6, #0285, AMAC, Inc. Westbrook, ME) and human IgG1 (clone JL512, #0280, AMAC, Inc. Westbrook, ME). Serum samples were serially diluted into the wells and the presence of specific immunoglobulins detected with affinity isolated alkaline phosphatase conjugated goat anti-human Ig (polyvalent) that had been pre-adsorbed to minimize cross-reactivity with mouse immunoglobulins. Table 3 and Fig. 28 show the results of an ELISA assay for the presence of human IgM and IgG1 in the serum of two animals that developed from embryos injected with the transgene insert of plasmid pHCl. All of the control non-transgenic mice tested negative for expression of human IgM and IgG1 by this assay. Mice from two lines containing the pIGM1 NotI insert (lines #6 and 15) express human IgM but not human IgG1. We tested mice from 6 lines that contain the pHCl insert and found that 4 of the lines (lines #26, 38, 57 and 122) express both human IgM and human IgG1, while mice from two of the lines (lines #19 and 21) do not express detectable levels of human immunoglobulins. The pHCl transgenic mice that did not express human immunoglobulins were so-called G₀ mice that developed directly from microinjected embryos and may have been mosaic for the presence of the transgene. Southern blot analysis indicates that many of these mice contain one or fewer copies of the transgene per cell. The

detection of human IgM in the serum of pIGM1 transgenics, and human IgM and IgG1 in pHCl transgenics, provides evidence that the transgene sequences function correctly in directing VDJ joining, transcription, and isotype switching. One of the animals (#18) was negative for the transgene by Southern blot analysis, and showed no detectable levels of human IgM or IgG1. The second animal (#38) contained approximately 5 copies of the transgene, as assayed by Southern blotting, and showed detectable levels of both human IgM and IgG1. The results of ELISA assays for 11 animals that developed from transgene injected embryos is summarized in the table below (Table 3).

TABLE 3

Detection of human IgM and IgG1 in the serum of transgenic animals by ELISA assay

<u>injected</u> <u>animal #</u>	<u>transgene</u>	<u>approximate</u> <u>transgene</u> <u>copies per cell</u>	<u>human IgM</u>	<u>human IgG1</u>
6	pIGM1	1	++	-
7	pIGM1	0	-	-
9	pIGM1	0	-	-
10	pIGM1	0	-	-
12	pIGM1	0	-	-
15	pIGM1	10	++	-
18	pHC1	0	-	-
19	pHC1	1	-	-
21	pHC1	<1	-	-
26	pHC1	2	++	+

38

pHC1

5

++

+

Table 3 shows a correlation between the presence of integrated transgene DNA and the presence of transgene encoded immunoglobulins in the serum. Two of the animals that were found to contain the pHC1 transgene did not express detectable levels of human immunoglobulins. These were both low copy animals and may not have contained complete copies of the transgenes, or the animals may have been genetic mosaics (indicated by the <1 copy per cell estimated for animal #21), and the transgene containing cells may not have populated the hematopoietic lineage. Alternatively, the transgenes may have integrated into genomic locations that are not conducive to their expression. The detection of human IgM in the serum of pIGM1 transgenics, and human IgM and IgG1 in pHC1 transgenics, indicates that the transgene sequences function correctly in directing VDJ joining, transcription, and isotype switching.

F. cDNA clones

To assess the functionality of the pHC1 transgene in VDJ joining and class switching, as well the participation of the transgene encoded human B-cell receptor in B-cell development and allelic exclusion, the structure of immunoglobulin cDNA clones derived from transgenic mouse spleen mRNA were examined. The overall diversity of the transgene encoded heavy chains, focusing on D and J segment usage, N region addition, CDR3 length distribution, and the frequency of joints resulting in functional mRNA molecules was examined. Transcripts encoding IgM and IgG incorporating VH105 and VH251 were examined.

Polyadenylated RNA was isolated from an eleven week old male second generation line-57 pHC1 transgenic mouse. This RNA was used to synthesize oligo-dT primed single stranded cDNA. The resulting cDNA was then used as template for four individual PCR amplifications using the following four synthetic oligonucleotides as primers: VH251 specific oligo-149, cta gct cga gtc caa gga gtc tgt gcc gag gtg cag ctg

(g,a,t,c) (SEQ ID NO:82); VH105 specific o-150, gtt gct cga gtg aaa ggt gtc cag tgt gag gtg cag ctg (g,a,t,c) (SEQ ID NO:83); human gamma1 specific oligo-151, ggc gct cga gtt cca cga cac cgt cac cgg ttc (SEQ ID NO:84); and human mu specific oligo-152, cct gct cga ggc agc caa cgg cca cgc tgc tcg (SEQ ID NO:85). Reaction 1 used primers o-149 and o-151 to amplify VH251-gamma1 transcripts, reaction 2 used o-149 and o-152 to amplify VH251-mu transcripts, reaction 3 used o-150 and o-151 to amplify VH105-gamma1 transcripts, and reaction 4 used o-150 and o-152 to amplify VH105-mu transcripts. The resulting 0.5 kb PCR products were isolated from an agarose gel; the μ transcript products were more abundant than the γ transcript products, consistent with the corresponding ELISA data (Fig. 34). The PCR products were digested with XhoI and cloned into the plasmid pNN03. Double-stranded plasmid DNA was isolated from minipreps of nine clones from each of the four PCR amplifications and dideoxy sequencing reactions were performed. Two of the clones turned out to be deletions containing no D or J segments. These could not have been derived from normal RNA splicing products and are likely to have originated from deletions introduced during PCR amplification. One of the DNA samples turned out to be a mixture of two individual clones, and three additional clones did not produce readable DNA sequence (presumably because the DNA samples were not clean enough). The DNA sequences of the VDJ joints from the remaining 30 clones are compiled in Table 4. Each of the sequences are unique, indicating that no single pathway of gene rearrangement, or single clone of transgene expressing B-cells is dominant. The fact that no two sequences are alike is also an indication of the large diversity of immunoglobulins that can be expressed from a compact minilocus containing only 2 V segments, 10 D segments, and 6 J segments. Both of the V segments, all six of the J segments, and 7 of the 10 D segments that are included in the transgene are used in VDJ joints. In addition, both constant region genes (μ and γ) are incorporated into transcripts. The VH105 primer turned out not to be specific for VH105 in the reactions performed. Therefore many of the

clones from reactions 3 and 4 contained VH251 transcripts. Additionally, clones isolated from ligated reaction 3 PCR product turned out to encode IgM rather than IgG; however this may reflect contamination with PCR product from reaction 4 as the DNA was isolated on the same gel. An analogous experiment, in which immunoglobulin heavy chain sequences were amplified from adult human peripheral blood lymphocytes (PBL), and the DNA sequence of the VDJ joints determined, was recently reported by Yamada et al. (J. Exp. Med. 173:395-407 (1991), which is incorporated herein by reference). We compared the data from human PBL with our data from the pHCl transgenic mouse.

TABLE 4 (SEQ ID NOS:86-115)

V		J		C		SEQ ID
n-D-n		n-D-n		n-D-n		
1	VH251 DHQ52 J3 Y 1	TACTGTGCGAGA	CGGCTACTGGGGTTGAT	GCCTTTTGATATCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	CCTCCACCAAG	[86]
2	VH251 DN1 J4 Y1	TACTGTGCGAGA	CACCGTATAGCAGCAGCTGG	CTTTTGACTACTGGGGCCAGGGAACCCCTGGTCAACGGTCTCTTCAG	CCTCCACCAAG	[87]
3	VH251 DP J6 Y 1	TACTGTGCGAGA	T	ATTACTACTACTACTAGGATATGAGAGCTCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	CCTCCACCAAG	[88]
4	VH251 DXP'1 J6 Y 1	TACTGTGCGAGA	CATTAGCATATTTTGACTGGTC	CTACTACTACTATGAGAGCTCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	CCTCCACCAAG	[89]
5	VH251 DXP'1 J4 Y 1	TACTGTGCGAGA	CGAGGTACTACTGGTTCGGGCAAGTTATTAAGCT	CTTTTGACTACTGGGGCCAGGGAACCCCTGGTCAACGGTCTCTTCAG	CCTCCACCAAG	[90]
6	VH251 DP J3 Y 1	TACTGTGCGAGA	CGGGGGGTGTCTGAT	GCCTTTTGATATCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	CCTCCACCAAG	[91]
7	VH251 DHQ52 J3 μ	TACTGTGCGAGA	GCMAATGGC	GCCTTTTGATATCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[92]
8	VH251 DHQ52 J6 μ	TACTGTGCGAGA	TGGGCTAACTGGGGATC	CTACTACTACTAGGATATGAGAGCTCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	GGAGTGCATTC	[93]
9	VH251 -- J1 μ	TACTGTGCGAGA		TACTTCCAGCATGGGGCCAGGGCACCCCTGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[94]
10	VH251 DLR2 J4 μ	TACTGTGCGAGA	CAGGTAGCTAACTCT	TTTGACTACTGGGGCCAGGGAACCCCTGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[95]
11	VH251 DXP'1 J4 μ	TACTGTGCGAGA	CAAAATTACTATGGTTGGGGAGTTCC	CTTTTGACTACTGGGGCCAGGGAACCCCTGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[96]
12	VH251 DP J1 μ	TACTGTGCGAGA	C	AAATACTTCCAGCACTGGGGCCAGGGAACCCCTGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[97]
13	VH251 DHQ52 J6 μ	TACTGTGCGAGA	CAAACTGGGG	ACTACTACTACTAGGATATGAGAGCTCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[98]
14	VH251 DXP'1 J6 μ	TACTGTGCGAGA	CATTACTATGGTTCGGGAGTTATG	ACTACTACTACTACTAGGATATGAGAGCTCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[99]
15	VH251 DXP'1 J4 Y1	TACTGTGCGAGA	CAGGGG	TGGGGCCAGGGGAACCCCTGGTCAACGGTCTCTTCAG	CCTCCACCAAG	[100]
16	VH105 DXP'1 J5 μ	TACTGTGCGAGA	TTCTGGGAG	ACTGGTTGACCCCTGGGGCCAGGGGAACCCCTGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[101]
17	VH251 DXP'1 J4 Y 1	TACTGTGCGAGA	CGAGGTACTACTGGTTCGGGAGTTATTAAGCT	CTTTTGACTACTGGGGCCAGGGAACCCCTGGTCAACGGTCTCTTCAG	CCTCCACCAAG	[102]
18	VH251 DHQ52 J4 Y 1	TACTGTGCGAGA	CAAACTGGGGAGA	GACTACTGGGGCCAGGGGAACCCCTGGTCAACGGTCTCTTCAG	CCTCCACCAAG	[103]
19	VH251 DK1 J6 Y 1	TACTGTGCGAGA	GGATATAGTGGCTACGATA	ACTACTACTACTAGGATATGAGAGCTCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	CCTCCACCAAG	[104]
20	VH251 DHQ52 J4 μ	TACTGTGCGAGA	CAAACTGGGGAGG	ACTACTTTTGACTACTGGGGCCAGGGGAACCCCTGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[105]
21	VH251 DK1 J2 Y 1	TACTGTGCGAGA	TATAGTGGCTACGATTAC	CTACTGGTACTTTGATCTCTGGGGCCAGGGGAACCCCTGGTCAACGGTCTCTTCAG	CCTCCACCAAG	[106]
22	VH251 DIR2 J6 Y 1	TACTGTGCGAGA	GCATCCCTCCCTCTCTTTG	ACTACTACTAGGATATGAGAGCTCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	CCTCCACCAAG	[107]
23	VH251 DIR2 J4 μ	TACTGTGCGAGA	CGGGGTGGGG	TTTGACTACTGGGGCCAGGGGAACCCCTGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[108]
24	VH105 DP J6 μ	TACTGTGCGAGA	CCGGTCAAACT	TTTACTACTACTACTAGGATATGAGAGCTCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[109]
25	VH105 DXP1 J4 μ	TACTGTGCGAGA	GATATTTTGACTGGTTAAG	TGACTACTGGGGCCAGGGGAACCCCTGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[110]
26	VH251 DN1 J3 μ	TACTGTGCGAGA	CATGGTATACAGCAGCTGGTAC	TGCTTTTGATATCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[111]
27	VH105 DHQ52 J3 μ	TACTGTGCGAGA	TCAACTGGGGTTG	ATGGTTTTTGATATCTGGGGCCAAAGGACAAATGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[112]
28	VH251 DN1 J4 μ	TACTGTGCG	GAATATGACAGCAGCTGCC	CTACTTTTGACTACTGGGGCCAGGGGAACCCCTGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[113]
29	VH105 DN1 J4 μ	TACTGTGCG	TGTATAGCAGCAGCTGGTAAAGGAACGG	CTACTGGGGCCAGGGGAACCCCTGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[114]
30	VH251 DHQ52 J4 μ	TACTGTGCGAGA	CAAACTGGGG	TGACTACTGGGGCCAGGGGAACCCCTGGTCAACGGTCTCTTCAG	GGAGTGCATCC	[115]

G. J segment choice

Table 5 compared the distribution of J segments incorporated into pHCl transgene encoded transcripts to J segments found in adult human PBL immunoglobulin transcripts. The distribution profiles are very similar, J4 is the dominant segment in both systems, followed by J6. J2 is the least common segment in human PBL and the transgenic animal.

TABLE 5
J. Segment Choice

<u>J. Segment</u>	Percent Usage ($\pm 3\%$)	
	<u>HC1 transgenic</u>	<u>Human PBL</u>
J1	7	1
J2	3	<1
J3	17	9
J4	44	53
J5	3	15
J6	<u>26</u>	<u>22</u>
	100%	100%

H. D segment choice

49% (40 of 82) of the clones analyzed by Yamada et al. incorporated D segments that are included in the pHCl transgene. An additional 11 clones contained sequences that were not assigned by the authors to any of the known D segments. Two of these 11 unassigned clones appear to be derived from an inversion of the DIR2 segments which is included in the pHCl construct. This mechanism, which was predicted by Ichihara et al. (EMBO J. 7:4141 (1988)) and observed by Sanz (J. Immunol. 147:1720-1729 (1991)), was not considered by Yamada et al. (J. Exp. Med. 173:395-407 (1991)). Table 5 is a comparison of the D segment distribution for the pHCl transgenic mouse and that observed for human PBL transcripts by Yamada et al. The data of Yamada et al. was recompiled to include DIR2 use, and to exclude D segments that are not in the pHCl transgene. Table 6 demonstrates that the distribution of D segment incorporation is very similar in the transgenic mouse and in human PBL. The two dominant human D

segments, DXP'1 and DN1, are also found with high frequency in the transgenic mouse. The most dramatic dissimilarity between the two distributions is the high frequency of DHQ52 in the transgenic mouse as compared to the human. The high frequency of DHQ52 is reminiscent of the D segment distribution in the human fetal liver. Sanz has observed that 14% of the heavy chain transcripts contained DHQ52 sequences. If D segments not found in pHCl are excluded from the analysis, 31% of the fetal transcripts analyzed by Sanz contain DHQ52. This is comparable to the 27% that we observe in the pHCl transgenic mouse.

TABLE 6

D. Segment	D Segment Choice	
	Percent Usage (\pm 3%)	
	<u>HC1 transgenic</u>	<u>Human PBL</u>
DLR1	<1	<1
DXP1	3	6
DXP'1	25	19
DA1	<1	12
DK1	7	12
DN1	12	22
DIR2	7	4
DM2	<1	2
DLR2	3	4
DHQ52	26	2
?	<u>17</u>	<u>17</u>
	100%	100%

I. Functionality of VDJ joints

Table 7 shows the predicted amino acid sequences of the VDJ regions from 30 clones that were analyzed from the pHCl transgenic. The translated sequences indicate that 23 of the 30 VDJ joints (77%) are in-frame with respect to the variable and J segments.

TABLE 7

Functionality 0 of V-D-J Joints (SEQ ID NOS:116-145)

			FR3	CDR3	FR4
1	VH251	DHQ52 J3 γ 1	YCAR	RLTGVDAFDI	WGQGTMTVTVSSASTK
2	VH251	DN1 J4 γ 1	YCAR	HRIAAAGFDY	WGQGTTLVTVSSASTK
3	VH251	D? J6 γ 1	YCAR	YYYYYGMDV	WGQGTTLVTVSSASTK
4	VH251	DXP'1J6 γ 1	YCAR	HYDILTGP TTTTVWTSGAKGPRSPSPQPPP	
5	VH251	DXP'1J4 γ 1	YCAR	RRYYGSGSYYNVTFDYWGQGTTLVTVSSASTK	
6	VH251	D? J3 γ 1	YCAR	RGVSDAFDI	WGQGTMTVTVSSASTK
7	VH251	DHQ52 J3 μ	YCAR	ATGAFDI	WGQGTMTVTVSSGSAS
8	VH251	DHQ52 J6 μ	YCARS	ANWGSYYYYYGMDVWGQGTTLVTVSSGSAS	
9	VH251	-- J1 μ	YCAR	YFQH	WGQGTTLVTVSSGSAS
10	VH251	DLR2 J4 μ	YCAR	HVANSFDY	WGQGTTLVTVSSGSAS
11	VH251	DXP'1 J4 μ	YCAR	QITMVRGV PFDY	WGQGTTLVTVSSGSAS
12	VH251	D? J1 μ	YCAR	QYFQH	WGQGTTLVTVSSGSAS
13	VH251	DHQ52 J6 μ	YCAR	QTGDYYYYYGMDVWGQGTTLVTVSSGSAS	
14	VH251	DXP'1 J6 μ	YCARHY	YGSYSYDY YYGMDVWGQGTTLVTVSSGSAS	
15	VH251	DXP'1 J4 γ 1	YCVR	QGVGPGNPGHRLLSLHQ	
16	VH105	DXP'1 J5 μ	YCAR	FWETGSTPGAREPWSPSPQGVH	
17	VH251	DXP'1 J4 γ 1	YCAR	RRYYGSGY YNVFDYWGQGTTLVIVSSASTK	
18	VH251	DHQ52 J4 γ 1	YCAR	QTWGGDY	WGQGTTLVTVSSASTK
19	VH251	DPQ J6 γ 1	YCAR	GYSGYDNY YYG I H V WGQGTTLVTVSSASTK	
20	VH251	DHQ52 J4 μ	YCAR	QTGEDYFDY	WGQGTTLVTVSSGSAS
21	VH251	DK1 J2 μ	YCAR	YSGYDYLLVLRSLGPWHPGHCLLSLHR	
22	VH251	DIR2 J6 γ 1	YCAR	ASLPSFDY YGMDV	WGQGTTLVTVSSASTK
23	VH251	DIR2 J4 μ	YCAR	RGGGLTTGAREPWSPSPQGVH	
24	VH105	D? J6 μ	YCV P	VETLLLLLRYGRLGPRDHGHRLLRECI	
25	VH105	DXP1 J4 μ	YCVR	DILTGXP DY	WGQGTTLVTVSSGSAS
26	VH251	DN1 J3 μ	YCAR	HGIAAAGTAFDI	WGQGTMTVTVSSGSAS
27	VH105	DHQ52 J3 μ	YCVR	STGVDAFDI	WGQGTMTVTVSSGSAS
28	VH251	DN1 J4 μ	YCAE	IAAAALLXLLGPGNPGHRLLECI	
29	VH105	DN1 J4 μ	YCV C	IAAAGKNGY	WGQGTTLVTVSSGSAS
30	VH251	DHQS2J4 μ	YCAR	QNWGDY	WGQGTTLVTVSSGSAS

J. CDR3 length distribution

Table 8 compared the length of the CDR3 peptides from transcripts with in-frame VDJ joints in the pHCl transgenic mouse to those in human PBL. Again the human PBL data comes from Yamada et al. The profiles are similar with the transgenic profile skewed slightly toward smaller CDR3 peptides than observed from human PBL. The average length of CDR3 in the transgenic mouse is 10.3 amino acids. This is substantially the same as the average size reported for authentic human CDR3 peptides by Sanz (J. Immunol. 147:1720-1729 (1991)).

TABLE 8
CDR3 Length Distribution

<u>#amino acids in CDR3</u>	<u>Percent Occurrence (\pm 3%)</u>	
	<u>HC1 transgenic</u>	<u>Human PBL</u>
3-8	26	14
9-12	48	41
13-18	26	37
19-23	<1	7
>23	<1	1
	100%	100%

EXAMPLE 13Rearranged Heavy Chain TransgenesA. Isolation of Rearranged Human Heavy Chain VDJ segments.

Two human leukocyte genomic DNA libraries cloned into the phage vector λ EMBL3/SP6/T7 (Clontech Laboratories, Inc., Palo Alto, CA) are screened with a 1 kb PacI/HindIII fragment of λ 1.3 containing the human heavy chain J- μ intronic enhancer. Positive clones are tested for hybridization with a mixture of the following V_H specific oligonucleotides:

oligo-7 5' - tca gtg aag gtt tcc tgc aag gca tct
gga tac acc ttc acc - 3' (SEQ ID NO:146)

oligo-8 5' - tcc ctg aga ctc tcc tgt gca gcc tct
gga ttc acc ttc agt - 3' (SEQ ID NO:147)

Clones that hybridized with both V and J- μ probes are isolated and the DNA sequence of the rearranged VDJ segment determined.

B. Construction of rearranged human heavy chain transgenes

Fragments containing functional VJ segments (open reading frame and splice signals) are subcloned into the plasmid vector pSP72 such that the plasmid derived XhoI site is adjacent to the 5' end of the insert sequence. A subclone containing a functional VDJ segment is digested with XhoI and PacI (PacI, a rare-cutting enzyme, recognizes a site near the J-m intronic enhancer), and the insert cloned into XhoI/PacI digested pH2 to generate a transgene construct with a functional VDJ segment, the J- μ intronic enhancer, the μ switch element, the μ constant region coding exons, and the γ 1 constant region, including the sterile transcript associated sequences, the γ 1 switch, and the coding exons. This transgene construct is excised with NotI and microinjected into the pronuclei of mouse embryos to generate transgenic animals as described above.

EXAMPLE 14

Light Chain Transgenes

A. Construction of Plasmid vectors

1. Plasmid vector pGP1c

Plasmid vector pGP1a is digested with NotI and the following oligonucleotides ligated in:

oligo-81 5' - ggc cgc atc ccg ggt ctc gag gtc gac
aag ctt tcg agg atc cgc - 3' (SEQ ID NO:148)

oligo-82 5' - ggc cgc gga tcc tcg aaa gct tgt cga
cct cga gac ccg gga tgc - 3' (SEQ ID NO:149)

The resulting plasmid, pGP1c, contains a polylinker with XmaI, XhoI, SalI, HindIII, and BamHI restriction sites flanked by NotI sites.

2. Plasmid vector pGP1d

Plasmid vector pGP1a is digested with NotI and the following oligonucleotides ligated in:

oligo-87 5' - ggc cgc tgt cga caa gct tat cga tgg
atc ctc gag tgc - 3' (SEQ ID NO:150)

oligo-88 5' - ggc cgc act cga gga tcc atc gat aag
ctt gtc gac agc - 3' (SEQ ID NO:151)

The resulting plasmid, pGP1d, contains a polylinker with SalI, HindIII, ClaI, BamHI, and XhoI restriction sites flanked by NotI sites.

B. Isolation of Jk and Ck clones

A human placental genomic DNA library cloned into the phage vector λEMBL3/SP6/T7 (Clonetech Laboratories, Inc., Palo Alto, CA) was screened with the human kappa light chain J region specific oligonucleotide:

oligo-36 5' - cac ctt cgg cca agg gac acg act gga
gat taa acg taa gca -3' (SEQ ID NO:152)

and the phage clones 136.2 and 136.5 isolated. A 7.4 kb XhoI fragment that includes the Jk1 segment was isolated from 136.2 and subcloned into the plasmid pNNO3 to generate the plasmid clone p36.2. A neighboring 13 kb XhoI fragment that includes Jk segments 2 through 5 together with the Ck gene segment was isolated from phage clone 136.5 and subcloned into the plasmid pNNO3 to generate the plasmid clone p36.5. Together these two clones span the region beginning 7.2 kb upstream of Jk1 and ending 9 kb downstream of Ck.

C. Construction of rearranged light chain transgenes

1. pCK1, a Ck vector for expressing rearranged variable segments

The 13 kb XhoI insert of plasmid clone p36.5 containing the Ck gene, together with 9 kb of downstream sequences, is cloned into the SalI site of plasmid vector pGP1c with the 5' end of the insert adjacent to the plasmid XhoI site. The resulting clone, pCK1 can accept cloned fragments containing rearranged VJk segments into the unique 5' XhoI site. The transgene can then be excised with NotI and purified from vector sequences by gel electrophoresis. The resulting transgene construct will contain the human J-Ck intronic enhancer and may contain the human 3' κ enhancer.

2. pCK2, a Ck vector with heavy chain enhancers for expressing rearranged variable segments

A 0.9 kb XbaI fragment of mouse genomic DNA containing the mouse heavy chain J- μ intronic enhancer (J. Banerji et al., Cell 33:729-740 (1983)) was subcloned into pUC18 to generate the plasmid pJH22.1. This plasmid was linearized with SphI and the ends filled in with Klenow enzyme. The Klenow treated DNA was then digested with HindIII and a 1.4 kb MluI/HindIII fragment of phage clone λ 1.3 (previous example), containing the human heavy chain J- μ intronic enhancer (Hayday et al., Nature 307:334-340 (1984)), to it. The resulting plasmid, pMHE1, consists of the mouse and human heavy chain J- μ intronic enhancers ligated together into pUC18 such that they are excised on a single BamHI/HindIII fragment. This 2.3 kb fragment is isolated and cloned into pGP1c to generate pMHE2. pMHE2 is digested with SalI and the 13 kb XhoI insert of p36.5 cloned in. The resulting plasmid, pCK2, is identical to pCK1, except that the mouse and human heavy chain J- μ intronic enhancers are fused to the 3' end of the transgene insert. To modulate expression of the final transgene, analogous constructs can be generated with different enhancers, i.e. the mouse or rat 3' kappa or heavy chain enhancer (Meyer and Neuberger, EMBO J.,

8:1959-1964 (1989); Petterson et al., Nature, 344:165-168 (1990)).

3. Isolation of rearranged kappa light chain variable segments

Two human leukocyte genomic DNA libraries cloned into the phage vector λ EMBL3/SP6/T7 (Clonetech Laboratories, Inc., Palo Alto, CA) were screened with the human kappa light chain J region containing 3.5 kb XhoI/SmaI fragment of p36.5. Positive clones were tested for hybridization with the following Vk specific oligonucleotide:

oligo-65 5' - agg ttc agt ggc agt ggg tct ggg aca
gac ttc act ctc acc atc agc - 3' (SEQ ID NO:153)

Clones that hybridized with both V and J probes are isolated and the DNA sequence of the rearranged VJk segment determined.

4. Generation of transgenic mice containing rearranged human light chain constructs.

Fragments containing functional VJ segments (open reading frame and splice signals) are subcloned into the unique XhoI sites of vectors pCK1 and pCK2 to generate rearranged kappa light chain transgenes. The transgene constructs are isolated from vector sequences by digestion with NotI. Agarose gel purified insert is microinjected into mouse embryo pronuclei to generate transgenic animals. Animals expressing human kappa chain are bred with heavy chain minilocus containing transgenic animals to generate mice expressing fully human antibodies.

Because not all VJk combinations may be capable of forming stable heavy-light chain complexes with a broad spectrum of different heavy chain VDJ combinations, several different light chain transgene constructs are generated, each using a different rearranged VJk clone, and transgenic mice that result from these constructs are bred with heavy chain minilocus transgene expressing mice. Peripheral blood, spleen, and lymph node lymphocytes are isolated from double

transgenic (both heavy and light chain constructs) animals, stained with fluorescent antibodies specific for human and mouse heavy and light chain immunoglobulins (Pharmingen, San Diego, CA) and analyzed by flow cytometry using a FACScan analyzer (Becton Dickinson, San Jose, CA). Rearranged light chain transgenes constructs that result in the highest level of human heavy/light chain complexes on the surface of the highest number of B cells, and do not adversely affect the immune cell compartment (as assayed by flow cytometric analysis with B and T cell subset specific antibodies), are selected for the generation of human monoclonal antibodies.

D. Construction of unrearranged light chain minilocus transgenes

1. pJCK1, a Jk, Ck containing vector for constructing minilocus transgenes

The 13 kb Ck containing XhoI insert of p36.5 is treated with Klenow enzyme and cloned into HindIII digested, Klenow-treated, plasmid pGP1d. A plasmid clone is selected such that the 5' end of the insert is adjacent to the vector derived ClaI site. The resulting plasmid, p36.5-1d, is digested with ClaI and Klenow-treated. The Jk1 containing 7.4 kb XhoI insert of p36.2 is then Klenow-treated and cloned into the ClaI, Klenow-treated p36.5-1d. A clone is selected in which the p36.2 insert is in the same orientation as the p36.5 insert. This clone, pJCK1 (Fig. 34), contains the entire human Jk region and Ck, together with 7.2 kb of upstream sequences and 9 kb of downstream sequences. The insert also contains the human J-Ck intronic enhancer and may contain a human 3' κ enhancer. The insert is flanked by a unique 3' SalI site for the purpose of cloning additional 3' flanking sequences such as heavy chain or light chain enhancers. A unique XhoI site is located at the 5' end of the insert for the purpose of cloning in unrearranged Vk gene segments. The unique SalI and XhoI sites are in turn flanked by NotI sites that are used to isolate the completed transgene construct away from vector sequences.

2. Isolation of unrearranged Vk gene segments and generation of transgenic animals expressing human Ig light chain protein

The Vk specific oligonucleotide, oligo-65 (discussed above), is used to probe a human placental genomic DNA library cloned into the phage vector λ EMBL3/SP6/T7 (Clonetech Laboratories, Inc., Palo Alto, CA). Variable gene segments from the resulting clones are sequenced, and clones that appear functional are selected. Criteria for judging functionality include: open reading frames, intact splice acceptor and donor sequences, and intact recombination sequence. DNA fragments containing selected variable gene segments are cloned into the unique XhoI site of plasmid pJCK1 to generate minilocus constructs. The resulting clones are digested with NotI and the inserts isolated and injected into mouse embryo pronuclei to generate transgenic animals. The transgenes of these animals will undergo V to J joining in developing B-cells. Animals expressing human kappa chain are bred with heavy chain minilocus containing transgenic animals to generate mice expressing fully human antibodies.

EXAMPLE 15

Genomic Heavy Chain Human Ig Transgene

This Example describes the cloning of a human genomic heavy chain immunoglobulin transgene which is then introduced into the murine germline via microinjection into zygotes or integration in ES cells.

Nuclei are isolated from fresh human placental tissue as described by Marzluff, W.F., et al. (1985), Transcription and Translation: A Practical Approach, B.D. Hammes and S.J. Higgins, eds., pp. 89-129, IRL Press, Oxford). The isolated nuclei (or PBS washed human spermatocytes) are embedded in 0.5% low melting point agarose blocks and lysed with 1 mg/ml proteinase K in 500mM EDTA, 1% SDS for nuclei, or with 1mg/ml proteinase K in 500mM EDTA, 1% SDS, 10mM DTT for spermatocytes at 50°C for 18 hours. The proteinase K is inactivated by incubating the blocks in 40µg/ml PMSF in TE for 30 minutes at 50°C, and then washing extensively with TE. The

DNA is then digested in the agarose with the restriction enzyme NotI as described by M. Finney in Current Protocols in Molecular Biology (F. Ausubel et al., eds. John Wiley & Sons, Supp. 4, 1988, e.g., Section 2.5.1).

The NotI digested DNA is then fractionated by pulsed field gel electrophoresis as described by Anand et al., Nuc. Acids Res. 17:3425-3433 (1989). Fractions enriched for the NotI fragment are assayed by Southern hybridization to detect one or more of the sequences encoded by this fragment. Such sequences include the heavy chain D segments, J segments, and $\gamma 1$ constant regions together with representatives of all 6 V_H families (although this fragment is identified as 670 kb fragment from HeLa cells by Berman et al. (1988), supra., we have found it to be an 830 kb fragment from human placental and sperm DNA). Those fractions containing this NotI fragment are ligated into the NotI cloning site of the vector pYACNN as described (McCormick et al., Technique 2:65-71 (1990)). Plasmid pYACNN is prepared by digestion of pYACneo (Clontech) with EcoRI and ligation in the presence of the oligonucleotide 5' - AAT TGC GGC CGC - 3' (SEQ ID NO:25).

YAC clones containing the heavy chain NotI fragment are isolated as described by Traver et al., Proc. Natl. Acad. Sci. USA, 86:5898-5902 (1989). The cloned NotI insert is isolated from high molecular weight yeast DNA by pulse field gel electrophoresis as described by M. Finney, op. cit. The DNA is condensed by the addition of 1 mM spermine and microinjected directly into the nucleus of single cell embryos previously described. Alternatively, the DNA is isolated by pulsed field gel electrophoresis and introduced into ES cells by lipofection (Gnirke et al., EMBO J. 10:1629-1634 (1991)), or the YAC is introduced into ES cells by spheroplast fusion.

EXAMPLE 16

Discontinuous Genomic Heavy Chain Ig Transgene

An 85 kb SpeI fragment of human genomic DNA, containing V_H6 , D segments, J segments, the μ constant region and part of the γ constant region, has been isolated by YAC cloning essentially as described in Example 1. A YAC carrying

a fragment from the germline variable region, such as a 570 kb NotI fragment upstream of the 670-830 kb NotI fragment described above containing multiple copies of V₁ through V₅, is isolated as described. (Berman et al. (1988), supra. detected two 570 kb NotI fragments, each containing multiple V segments.) The two fragments are coinjected into the nucleus of a mouse single cell embryo as described in Example 1.

Typically, coinjection of two different DNA fragments result in the integration of both fragments at the same insertion site within the chromosome. Therefore, approximately 50% of the resulting transgenic animals that contain at least one copy of each of the two fragments will have the V segment fragment inserted upstream of the constant region containing fragment. Of these animals, about 50% will carry out V to DJ joining by DNA inversion and about 50% by deletion, depending on the orientation of the 570 kb NotI fragment relative to the position of the 85 kb SpeI fragment. DNA is isolated from resultant transgenic animals and those animals found to be containing both transgenes by Southern blot hybridization (specifically, those animals containing both multiple human V segments and human constant region genes) are tested for their ability to express human immunoglobulin molecules in accordance with standard techniques.

EXAMPLE 17

Identification of functionally rearranged variable region sequences in transgenic B cells

An antigen of interest is used to immunize (see Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor, New York (1988)) a mouse with the following genetic traits: homozygosity at the endogenous heavy chain locus for a deletion of J_H (Examples 10); hemizygous for a single copy of unrearranged human heavy chain minilocus transgene (examples 5 and 14); and hemizygous for a single copy of a rearranged human kappa light chain transgene (Examples 6 and 14).

Following the schedule of immunization, the spleen is removed, and spleen cells used to generate hybridomas.

Cells from an individual hybridoma clone that secretes antibodies reactive with the antigen of interest are used to prepare genomic DNA. A sample of the genomic DNA is digested with several different restriction enzymes that recognize unique six base pair sequences, and fractionated on an agarose gel. Southern blot hybridization is used to identify two DNA fragments in the 2-10 kb range, one of which contains the single copy of the rearranged human heavy chain VDJ sequences and one of which contains the single copy of the rearranged human light chain VJ sequence. These two fragments are size fractionated on agarose gel and cloned directly into pUC18. The cloned inserts are then subcloned respectively into heavy and light chain expression cassettes that contain constant region sequences.

The plasmid clone p γ el (Example 12) is used as a heavy chain expression cassette and rearranged VDJ sequences are cloned into the XhoI site. The plasmid clone pCK1 is used as a light chain expression cassette and rearranged VJ sequences are cloned into the XhoI site. The resulting clones are used together to transfect SP₀ cells to produce antibodies that react with the antigen of interest (Co. et al., Proc. Natl. Acad. Sci. USA 88:2869 (1991), which is incorporated herein by reference).

Alternatively, mRNA is isolated from the cloned hybridoma cells described above, and used to synthesize cDNA. The expressed human heavy and light chain VDJ and VJ sequence are then amplified by PCR and cloned (Larrick et al., Biol. Technology, 7:934-938 (1989)). After the nucleotide sequence of these clones has been determined, oligonucleotides are synthesized that encode the same polypeptides, and synthetic expression vectors generated as described by Queen et al., Proc. Natl. Acad. Sci. USA, 84:5454-5458 (1989).

Immunization of Transgenic Animals with Complex Antigens

The following experiment demonstrates that transgenic animals can be successfully immunized with complex antigens such as those on human red blood cells and respond

with kinetics that are similar to the response kinetics observed in normal mice.

Blood cells generally are suitable immunogens and comprise many different types of antigens on the surface of red and white blood cells.

Immunization with human blood

Tubes of human blood from a single donor were collected and used to immunize transgenic mice having functionally disrupted endogenous heavy chain loci ($J_H D$) and harboring a human heavy chain minigene construct (HC1); these mice are designated as line 112. Blood was washed and resuspended in 50 mls Hanks' and diluted to 1×10^8 cells/ml. 0.2 mls (2×10^7 cells) were then injected interperitoneally using a 28 gauge needle and 1 cc syringe. This immunization protocol was repeated approximately weekly for 6 weeks. Serum titers were monitored by taking blood from retro-orbital bleeds and collecting serum and later testing for specific antibody. A pre-immune bleed was also taken as a control. On the very last immunization, three days before these animals were sacrificed for serum and for hybridomas, a single immunization of 1×10^7 cells was given intravenously through the tail to enhance the production of hybridomas.

Table 9

Animals

	Mouse ID	Line	Sex	HC1-112	JHD
1	2343	112	M	+	++
2	2344	112	M	-	+
3	2345	112	F	-	+
4	2346	112	F	-	++
5	2347	112	F	-	++
6	2348	112	F	+	++
7	2349	112	F	-	+

1163

Mice # 2343 and 2348 have a desired phenotype: human heavy chain mini-gene transgenic on heavy chain knock-out background.

Generation of Hybridomas

Hybridomas were generated by fusing mouse spleen cells of approximately 16 week-old transgenic mice (Table 9) that had been immunized as described (supra) to a fusion partner consisting of the non-secreting HAT-sensitive myeloma cell line, X63 Ag8.653. Hybridoma clones were cultivated and hybridoma supernatants containing immunoglobulins having specific binding affinity for blood cell antigens were identified, for example, by flow cytometry.

Flow cytometry

Serum and hybridoma supernatants were tested using flow cytometry. Red blood cells from the donor were washed 4X in Hanks' balanced salt solution and 50,000 cells were placed in 1.1 ml polypropylene microtubes. Cells were incubated with antisera or supernatant from the hybridomas for 30 minutes on ice in staining media (1x RPMI 1640 media without phenol red or biotin (Irvine Scientific) 3% newborn calf serum, 0.1% Na azide). Controls consisted of littermate mice with other genotypes. Cells were then washed by centrifugation at 4°C in Sorvall RT600B for 5-10 minutes at 1000 rpm. Cells were washed two times and then antibody detected on the cell surface with a fluorescent developing reagent. Two monoclonal reagents were used to test. One was a FITC-labeled mouse anti-human μ heavy chain antibody (Pharmagen, San Diego, CA) and the other was a PE-labeled rat anti-mouse kappa light chain (Becton-Dickenson, San Jose, CA). Both of these reagents gave similar results. Whole blood (red blood cells and white blood cells) and white blood cells alone were used as target cells. Both sets gave positive results.

Serum of transgenic mice and littermate controls was incubated with either red blood cells from the donor, or white blood cells from another individual, washed and then developed with anti-human IgM FITC labeled antibody and analyzed in a

flow cytometer. Results showed that serum from mice that are transgenic for the human mini-gene locus (mice 2343 and 2348) show human IgM reactivity whereas all littermate animals (2344, 2345, 2346, 2347) do not. Normal mouse serum (NS) and phosphate buffer saline (PBS) were used as negative controls. Red blood cells were ungated and white blood cells were gated to include only lymphocytes. Lines are drawn on the x and y axis to provide a reference. Flow cytometry was performed on 100 supernatants from fusion 2348. Four supernatants showed positive reactivity for blood cell antigens.

EXAMPLE 18

Reduction of Endogenous Mouse Immunoglobulin Expression by Antisense RNA

A. Vector for Expression of Antisense Ig Sequences

1. Construction of the cloning vector pGP1h

The vector pGP1b (referred to in a previous example) is digested with XhoI and BamHI and ligated with the following oligonucleotides:

5'- gat cct cga gac cag gta cca gat ctt gtg aat tcg -3' (SEQ ID NO:154)

5'- tcg acg aat tca caa gat ctg gta cct ggt ctc gag -3' (SEQ ID NO:155)

to generate the plasmid pGP1h. This plasmid contains a polylinker that includes the following restriction sites: NotI, EcoRI, BglII, Asp718, XhoI, BamHI, HindIII, NotI.

Construction of pBCE1.

A 0.8 kb XbaI/BglII fragment of pVH251 (referred to in a previous example), that includes the promoter leader sequence exon, first intron, and part of the second exon of the human VH-V family immunoglobulin variable gene segment, was inserted into XbaI/BglII digested vector pNN03 to generate the plasmid pVH251.

The 2.2 kb BamHI/EcoRI DNA fragment that includes the coding exons of the human growth hormone gene (hGH;

Seeburg, (1982) DNA 1:239-249) is cloned into BglIII/EcoRI digested pGH1h. The resulting plasmid is digested with BamHI and the BamHI/BglIII of pVH251N is inserted in the same orientation as the hGH gene to generate the plasmid pVhgh.

A 0.9 kb XbaI fragment of mouse genomic DNA containing the mouse heavy chain J- μ intronic enhancer (Banerji et al., (1983) Cell 33:729-740) was subcloned into pUC18 to generate the plasmid pJH22.1. This plasmid was linearized with SphI and the ends filled in with klenow enzyme. The klenow treated DNA was then digested with HindIII and a 1.4 kb MluI(klenow)/HindIII fragment of phage clone λ 1.3 (previous example), containing the human heavy chain J- μ intronic enhancer (Hayday et al., (1984) Nature 307:334-340), to it. The resulting plasmid, pMHE1, consists of the mouse and human heavy chain J- μ intron enhancers ligated together into pUC18 such that they can be excised on a single BamHI/HindIII fragment.

The BamHI/HindIII fragment of pMHE1 is cloned into BamHI/HindIII cut pVhgh to generate the B-cell expression vector pBCE1. This vector, depicted in Fig. 36, contains unique XhoI and Asp718 cloning sites into which antisense DNA fragments can be cloned. The expression of these antisense sequences is driven by the upstream heavy chain promoter-enhancer combination the downstream hGH gene sequences provide polyadenylation sequences in addition to intron sequences that promote the expression of transgene constructs. Antisense transgene constructs generated from pBCE1 can be separated from vector sequences by digestion with NotI.

B. An IgM antisense transgene construct.

The following two oligonucleotides:

5'- cgc ggt acc gag agt cag tcc ttc cca aat gtc -3' (SEQ ID NO:156)

5'- cgc ctc gag aca gct gga atg ggc aca tgc aga -
3' (SEQ ID NO:157)

are used as primers for the amplification of mouse IgM constant region sequences by polymerase chain reaction (PCR) using mouse spleen cDNA as a substrate. The resulting 0.3 kb PCR product is digested with Asp718 and XhoI and cloned into Asp718/XhoI digested pBCE1 to generate the antisense transgene construct pMAS1. The purified NotI insert of pMAS1 is microinjected into the pronuclei of half day mouse embryos--alone or in combination with one or more other transgene constructs--to generate transgenic mice. This construct expresses an RNA transcript in B-cells that hybridizes with mouse IgM mRNA, thus down-regulating the expression of mouse IgM protein. Double transgenic mice containing pMAS1 and a human heavy chain transgene minilocus such as pHCl (generated either by coinjection of both constructs or by breeding of singly transgenic mice) will express the human transgene encoded Ig receptor on a higher percentage of B-cell than mice transgenic for the human heavy chain minilocus alone. The ratio of human to mouse Ig receptor expressing cells is due in part to competition between the two populations for factors and cells that promoter B-cell differentiation and expansion. Because the Ig receptor plays a key role in B-cell development, mouse Ig receptor expressing B-cells that express reduced levels of IgM on their surface (due to mouse Ig specific antisense down-regulation) during B-cell development will not compete as well as cells that express the human receptor.

C. An IgKappa antisense transgene construct.

The following two oligonucleotides:

5'- cgc ggt acc gct gat gct gca cca act gta tcc -3' (SEQ ID NO158)

5'- cgc ctc gag cta aca ctc att cct gtt gaa gct -3' (SEQ ID NO159)

are used as primers for the amplification of mouse IgKappa constant region sequences by polymerase chain reaction (PCR) using mouse spleen cDNA as a substrate. The resulting 0.3 kb

PCR product is digested with Asp718 and XhoI and cloned into Asp718/XhoI digested pBCE1 to generate the antisense transgene construct pKAS1. The purified NotI insert of pKAS1 is microinjected into the pronuclei of half day mouse embryos--alone or in combination with one or more other transgene constructs--to generate transgenic mice. This construct expresses an RNA transcript in B-cells that hybridizes with mouse IgK mRNA, thus down-regulating the expression of mouse IgK protein as described above for pMAS1.

EXAMPLE 19

This example demonstrates the successful immunization and immune response in a transgenic mouse of the present invention.

Immunization of Mice

Keyhole limpet hemocyanin conjugated with greater than 400 dinitrophenyl groups per molecule (Calbiochem, La Jolla, California) (KLH-DNP) was alum precipitated according to a previously published method (Practical Immunology, L. Hudson and F.C. Hay, Blackwell Scientific (Pubs.), p. 9, 1980). Four hundred µg of alum precipitated KLH-DNP along with 100 µg dimethyldioctadecyl Ammonium Bromide in 100 µL of phosphate buffered saline (PBS) was injected intraperitoneally into each mouse. Serum samples were collected six days later by retro-orbital sinus bleeding.

Analysis of Human Antibody Reactivity in Serum

Antibody reactivity and specificity were assessed using an indirect enzyme-linked immunosorbent assay (ELISA). Several target antigens were tested to analyze antibody induction by the immunogen. Keyhole limpet hemocyanin (Calbiochem) was used to identify reactivity against the protein component, bovine serum albumin-DNP for reactivity against the hapten and/or modified amino groups, and KLH-DNP for reactivity against the total immunogen. Human antibody binding to antigen was detected by enzyme conjugates specific for IgM and IgG sub-classes with no cross reactivity to mouse

immunoglobulin. Briefly, PVC microtiter plates were coated with antigen drying overnight at 37°C of 5 µg/mL protein in PBS. Serum samples diluted in PBS, 5% chicken serum, 0.5% Tween-20 were incubated in the wells for 1 hour at room temperature, followed by anti-human IgG Fc and IgG F(ab')-horseradish peroxidase or anti-human IgM Fc-horseradish peroxidase in the same diluent. After 1 hour at room temperature enzyme activity was assessed by addition of ABTS substrate (Sigma, St. Louis, Missouri) and read after 30 minutes at 415-490 nm.

Human Heavy Chain Participation in Immune Response in Transgenic Mice

Figures 37A-37D illustrate the response of three mouse littermates to immunization with KLH-DNP. Mouse number 1296 carried the human IgM and IgG unrearranged transgene and was homozygous for mouse Ig heavy chain knockout. Mouse number 1299 carried the transgene on a non-knockout background, while mouse 1301 inherited neither of these sets of genes. Mouse 1297, another littermate, carried the human transgene and was hemizygous with respect to mouse heavy chain knockout. It was included as a non-immunized control.

The results demonstrate that both human IgG and IgM responses were developed to the hapten in the context of conjugation to protein. Human IgM also developed to the KLH molecule, but no significant levels of human IgG were present at this time point. In pre-immunization serum samples from the same mice, titers of human antibodies to the same target antigens were insignificant.

EXAMPLE 20

This example demonstrates the successful immunization with a human antigen and immune response in a transgenic mouse of the present invention, and provides data demonstrating that nonrandom somatic mutation occurs in the variable region sequences of the human transgene.

Demonstration of antibody responses comprising human immunoglobulin heavy chains against a human glycoprotein antigen

Transgenic mice used for the experiment were homozygous for functionally disrupted murine immunoglobulin heavy chain loci produced by introduction of a transgene at the joining (J) region (supra) resulting in the absence of functional endogenous (murine) heavy chain production. The transgenic mice also harbored at least one complete unrearranged human heavy chain mini-locus transgene, (HC1, supra), which included a single functional V_H gene (V_H251), human μ constant region gene, and human γ 1 constant region gene. Transgenic mice shown to express human immunoglobulin transgene products (supra) were selected for immunization with a human antigen to demonstrate the capacity of the transgenic mice to make an immune response against a human antigen immunization. Three mice of the HC1-26 line and three mice of the HC1-57 line (supra) were injected with human antigen.

One hundred μ g of purified human carcinoembryonic antigen (CEA) insolubilized on alum was injected in complete Freund's adjuvant on Day 0, followed by further weekly injections of alum-precipitated CEA in incomplete Freund's adjuvant on Days 7, 14, 21, and 28. Serum samples were collected by retro-orbital bleeding on each day prior to injection of CEA. Equal volumes of serum were pooled from each of the three mice in each group for analysis.

Titres of human μ chain-containing immunoglobulin and human γ chain-containing immunoglobulin which bound to human CEA immobilized on microtitre wells were determined by ELISA assay. Results of the ELISA assays for human μ chain-containing immunoglobulins and human γ chain-containing immunoglobulins are shown in Figs. 38 and 39, respectively. Significant human μ chain Ig titres were detected for both lines by Day 7 and were observed to rise until about Day 21. For human γ chain Ig, significant titres were delayed, being evident first for line HC1-57 at Day 14, and later for line HC1-26 at Day 21. Titres for human γ chain Ig continued to show an increase over time during the course of the

experiment. The observed human μ chain Ig response, followed by a plateau, combined with a later developing γ chain response which continues to rise is characteristic of the pattern seen with affinity maturation. Analysis of Day 21 samples showed lack of reactivity to an unrelated antigen, keyhole limpet hemocyanin (KLC), indicating that the antibody response was directed against CEA in a specific manner.

These data indicate that animals transgenic for human unrearranged immunoglobulin gene loci: (1) can respond to a human antigen (e.g., the human glycoprotein, CEA), (2) can undergo isotype switching ("class switching") as exemplified by the observed μ to γ class switch, and (3) exhibit characteristics of affinity maturation in their humoral immune responses. In general, these data indicate: (1) the human Ig transgenic mice have the ability to induce heterologous antibody production in response to a defined antigen, (2) the capacity of a single transgene heavy chain variable region to respond to a defined antigen, (3) response kinetics over a time period typical of primary and secondary response development, (4) class switching of a transgene-encoded humoral immune response from IgM to IgG, and (5) the capacity of transgenic animal to produce human-sequence antibodies against a human antigen.

Demonstration of somatic mutation in a human heavy chain transgene minilocus.

Line HC1-57 transgenic mice, containing multiple copies of the HC1 transgene, were bred with immunoglobulin heavy chain deletion mice to obtain mice that contain the HC1 transgene and contain disruptions at both alleles of the endogenous mouse heavy chain (supra). These mice express human μ and γ heavy chains together with mouse κ and λ light chains (supra). One of these mice was hyperimmunized against human carcinoembryonic antigen by repeated intraperitoneal injections over the course of 1.5 months. This mouse was sacrificed and lymphoid cells isolated from the spleen, inguinal and mesenteric lymph nodes, and Peyer's patches. The cells were combined and total RNA

isolated. First strand cDNA was synthesized from the RNA and used as a template for PCR amplification with the following 2 oligonucleotide primers:

149 5' - cta gct cga gtc caa gga gtc tgt gcc gag gtg cag ctg
(g/a/t/c) - 3' (SEQ ID NO:82)

151 5' - ggc gct cga gtt cca cga cac cgt cac cgg ttc - 3' (SEQ
ID NO:84)

These primers specifically amplify VH251/gamma1 cDNA sequences. The amplified sequences were digested with XhoI and cloned into the vector pNN03. DNA sequence from the inserts of 23 random clones is shown in Fig. 40; sequence variations from germline sequence are indicated, dots indicate sequence is identical to germline. Comparison of the cDNA sequences with the germline sequence of the VH251 transgene reveals that 3 of the clones are completely unmutated, while the other 20 clones contain somatic mutations. One of the 3 non-mutated sequences is derived from an out-of-frame VDJ joint. Observed somatic mutations at specific positions of occur at similar frequencies and in similar distribution patterns to those observed in human lymphocytes (Cai et al. (1992) J. Exp. Med. 176: 1073, incorporated herein by reference). The overall frequency of somatic mutations is approximately 1%; however, the frequency goes up to about 5% within CDR1, indicating selection for amino acid changes that affect antigen binding. This demonstrates antigen driven affinity maturation of the human heavy chain sequences.

EXAMPLE 21

This example demonstrates the successful formation of a transgene by co-introduction of two separate polynucleotides which recombine to form a complete human light chain minilocus transgene.

Generation of an unrearranged light chain minilocus transgene by co-injection of two overlapping DNA fragments

1. Isolation of unrearranged functional V_k gene segments vk65.3, vk65.5, vk65.8 and vk65.15

The V_k specific oligonucleotide, oligo-65 (5'-agg ttc agt ggc agt ggg tct ggg aca gac ttc act ctc acc atc agc-3'; SEQ ID NO:153), was used to probe a human placental genomic DNA library cloned into the phage vector λEMBL3/SP6/T7 (Clonetech Laboratories, Inc., Palo Alto, CA). DNA fragments containing V_k segments from positive phage clones were subcloned into plasmid vectors. Variable gene segments from the resulting clones are sequenced, and clones that appear functional were selected. Criteria for judging functionality include: open reading frames, intact splice acceptor and donor sequences, and intact recombination sequence. DNA sequences of 4 functional V_k gene segments (vk65.3, vk65.5, vk65.8, and vk65.15) from 4 different plasmid clones isolated by this procedure are shown in Figs. 41-44. The four plasmid clones, p65.3f, p65.5g1, p65.8, and p65.15f, are described below.

(1 a) p65.3f

A 3 kb Xba fragment of phage clone λ65.3 was subcloned into pUC19 so that the vector derived SalI site was proximal to the 3' end of the insert and the vector derived BamHI site 5'. The 3 kb BamHI/SalI insert of this clone was subcloned into pGP1f to generate p65.3f.

(1 b) p65.5g1

A 6.8 kb EcoRI fragment of phage clone λ65.5 was subcloned into pGP1f so that the vector derived XhoI site is proximal to the 5' end of the insert and the vector derived SalI site 3'. The resulting plasmid is designated p65.5g1.

(1 c) p65.8

A 6.5 kb HindIII fragment of phage clone λ65.8 was cloned into pSP72 to generate p65.8.

(1 d) p65.15f

A 10 kb EcoRI fragment of phage clone λ 65.16 was subcloned into pUC18 to generate the plasmid p65.15.3. The V_k gene segment within the plasmid insert was mapped to a 4.6 kb EcoRI/HindIII subfragment, which was cloned into pGP1f. The resulting clone, p65.15f, has unique XhoI and SalI sites located at the respective 5' and 3' ends of the insert.

2. pKV4

The XhoI/SalI insert of p65.8 was cloned into the XhoI site of p65.15f to generate the plasmid pKV2. The XhoI/SalI insert of p65.5g1 was cloned into the XhoI site of pKV2 to generate pKV3. The XhoI/SalI insert of pKV3 was cloned into the XhoI site of p65.3f to generate the plasmid pKV4. This plasmid contains a single 21 kb XhoI/SalI insert that includes 4 functional V_k gene segments. The entire insert can also be excised with NotI.

3. pKC1B

(3 a) pKcor

Two XhoI fragments derived from human genomic DNA phage λ clones were subcloned into plasmid vectors. The first, a 13 kb J_k2 - J_k5 / C_k containing fragment, was treated with Klenow enzyme and cloned into HindIII digested, Klenow treated, plasmid pGP1d. A plasmid clone (pK-31) was selected such that the 5' end of the insert is adjacent to the vector derived ClaI site. The second XhoI fragment, a 7.4 kb piece of DNA containing J_k1 was cloned into XhoI/SalI-digested pSP72, such that the 3' insert XhoI site was destroyed by ligation to the vector SalI site. The resulting clone, p36.2s, includes an insert derived ClaI site 4.5 kb upstream of J_k1 and a polylinker derived ClaI site downstream in place of the naturally occurring XhoI site between J_k1 and J_k2 . This clone was digested with ClaI to release a 4.7 kb fragment which was cloned into ClaI digested pK-31 in the correct 5' to 3' orientation to generate a plasmid containing all 5 human J_k segments, the human intronic enhancer human C_k , 4.5 kb of 5' flanking sequence, and 9 kb of 3' flanking sequence. This

plasmid, pKcor, includes unique flanking XhoI and SalI sites on the respective 5' and 3' sides of the insert.

(3 b) pKcorB

A 4 kb BamHI fragment containing the human 3' kappa enhancer (Judde, J.-G. and Max, E.E. (1992) Mol. Cell. Biol. 12: 5206, incorporated herein by reference) was cloned into pGP1f such that the 5' end is proximal to the vector XhoI site. The resulting plasmid, p24Bf, was cut with XhoI and the 17.7 kb XhoI/SalI fragment of pKcor cloned into it in the same orientation as the enhancer fragment. The resulting plasmid, pKcorB, includes unique XhoI and SalI sites at the 5' and 3' ends of the insert respectively.

(3 c) pKC1B

The XhoI/SalI insert of pKcorB was cloned into the SalI site of p65.3f to generate the light-chain minilocus-transgene plasmid pKC1B. This plasmid includes a single functional human V_κ segment, all 5 human J_κ segments, the human intronic enhancer, human C_κ, and the human 3' kappa enhancer. The entire 25 kb insert can be isolated by NotI digestion.

4. Co4

The two NotI inserts from plasmids pKV4 and pKC1B were mixed at a concentration of 2.5 µg/ml each in microinjection buffer, and co-injected into the pronuclei of half day mouse embryos as described in previous examples. Resulting transgenic animals contain transgene inserts (designated Co4, product of the recombination shown in Fig. 45) in which the two fragments co-integrated. The 3' 3 kb of the pKV4 insert and the 5' 3 kb of the pKC1B insert are identical. Some of the integration events will represent homologous recombinations between the two fragments over the 3 kb of shared sequence. The Co4 locus will direct the expression of a repertoire of human sequence light chains in a transgenic mouse.

This example demonstrates the successful production of a murine hybridoma clone secreting a monoclonal antibody reactive with a specific immunogen, wherein the monoclonal antibody comprises a human immunoglobulin chain encoded by a human Ig transgene.

Generation of Monoclonal Antibodies Incorporating Human Heavy Chain Transgene Product

1. Immunization of Mouse Harboring Human Heavy Chain Transgene

A mouse containing a human heavy chain encoding transgene and homozygous for knockout (i.e., functional disruption) of the endogenous heavy chain locus (see, EXAMPLE 20, *supra*) was immunized with purified human CEA, and spleen cells were subsequently harvested after a suitable immune response period. The murine spleen cells were fused with mouse myeloma cells to generate hybridomas using conventional techniques (see, Kohler and Milstein, Eur. J. Immunol., 6:511-519 (1976); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor, New York (1988)). The mouse used for immunization contained a human unrearranged heavy chain minilocus transgene which comprised a single functional V_H gene (V_{H251}), human D and J segments, human μ constant region, and human γ 1 constant region genes. The transgenic line from which it originated was designated HC1-57 (*supra*).

One hundred μ g of purified human carcinoembryonic antigen (CEA) (Cyrstal Chem, Chicago, IL or Scripps Labs, San Diego, CA) insolubilized on alum was injected in complete Freund's adjuvant on Day 0, followed by further weekly injections of alum-precipitated CEA in incomplete Freund's adjuvant on Days 7, 14, 21, and 28. An additional 20 μ g of soluble CEA was administered intravenously on Day 83, followed by 50 μ g alum-precipitated CEA in incomplete Freund's adjuvant on Day 92. Human heavy chain responses to CEA were confirmed in serum samples prior to fusion of spleen cells with myeloma cells. The animal was sacrificed on Day 95, the spleen removed and fused with P3X63-Ag8.653 mouse myeloma cells (ATCC CRL 1580, American Type Culture Collection, Rockville, MD) using polyethylene glycol. Two weeks later, supernates from

fusion wells were screened for the presence of antibodies specifically reactive with CEA, and which contained human heavy chain μ or γ constant region epitopes by ELISA. Briefly, purified human CEA was coated onto PVC microtitre plates at 2.5 $\mu\text{g/ml}$, and incubate with culture supernate diluted 1:4 or 1:5 in PBS, 0.5% Tween-20, 5% chicken serum. Plates were washed, followed by addition of horseradish peroxidase-conjugated goat antiserum specific for human IgG Fc or rabbit antiserum specific for human IgM Fc5Mu (Jackson ImmunoResearch, West Grove, PA). Presence of conjugate bound to captured antibody was determined, after further washing, by the addition of ABTS substrate. Two independent fusion wells were found to contain antibody with substantial binding to CEA. After cloning, both hybridomas were found to be positive for the presence of human μ chain and murine κ chain by ELISA. No mouse IgG or IgM were detected using similar assays.

Subcloning of the two independent parent hybridomas resulted in two clones, designated 92-09A-4F7-A5-2 and 92-09A-1D7-1-7-1. Both lines were deposited with the ATCC Patent Culture Depository under the Budapest Treaty and were assigned ATCC Designation HB 11307 and HB 11308, respectively. Culture supernatants from these cell lines were assessed for specificity by testing for reactivity to several purified target proteins using ELISA. As shown in Fig. 46, ELISA assays for determining the reactivity of the monoclonal antibodies to various antigens demonstrate that only CEA and the CEA-related antigen NCA-2 show significant reactivity, indicating the development of a restricted reactivity for the variable regions of the heterohybrid immunoglobulin molecules.

EXAMPLE 23

This example demonstrates that a rearranged human VDJ gene encoded by a human Ig minilocus transgene may be transcribed as a transcript which includes an endogenous Ig constant region gene, for example by the mechanism of trans-switching, to encode a chimeric human/mouse Ig chain.

Identification of Trans-Switch Transcripts Encoding Chimeric Human-Mouse Heavy Chains

RNA was isolated from a hyperimmunized HC1 line 57 transgenic mouse homozygous for the endogenous heavy chain J segment deletion (*supra*). cDNA was synthesized according to Taylor et al. (1993) Nucleic Acids Res. 20: 6287, incorporated herein by reference, and amplified by PCR using the following two primers:

o-149 (human V_{H251}):

5'-CTA GCT CGA GTC CAA GGA GTC TGT GCC GAG GTG CAG CTG (G,A,T,C)-3' (SEQ ID NO:82)

o-249 (mouse gamma):

5'-GGC GCT CGA GCT GGA CAG GG(A/C) TCC A(G/T)A GTT CCA-3' (SEQ ID NO:160)

Oligonucleotide o-149 is specific for the HC1-encoded variable gene segment V_{H251}, while o-249 hybridizes to both mouse and human gamma sequences with the following order of specificities:

mouse γ 1 = mouse γ 2b = mouse γ 3 > mouse γ 2a >> human γ 1.

DNA sequences from 10 randomly chosen clones generated from the PCR products was determined and is shown in Fig. 47. Two clones comprised human VDJ and mouse γ 1; four clones comprised human VDJ and mouse γ 2b; and four clones comprised human VDJ and mouse γ 3. These results indicate that in a fraction of the transgenic B cells, the transgene-encoded human VDJ recombined into the endogenous murine heavy chain locus by class switching or an analogous recombination.

EXAMPLE 24

This example describes a method for screening a pool of hybridomas to discriminate clones which encode chimeric human/mouse Ig chains from clones which encode and express a human Ig chain. For example, in a pool of hybridoma clones made from a transgenic mouse comprising a human Ig heavy chain transgene and homozygous for a J region-disrupted endogenous heavy chain locus, hybridoma clones encoding trans-switched human VDJ-murine constant region heavy chains may be

identified and separated from hybridoma clones expressing human VDJ-human constant region heavy chains.

Screening Hybridomas to Eliminate Chimeric Ig Chains

The screening process involves two stages, which may be conducted singly or optionally in combination: (1) a preliminary ELISA-based screen, and (2) a secondary molecular characterization of candidate hybridomas. Preferably, a preliminary ELISA-based screen is used for initial identification of candidate hybridomas which express a human VDJ region and a human constant region.

Hybridomas that show positive reactivity with the antigen (e.g., the immunogen used to elicit the antibody response in the transgenic mouse) are tested using a panel of monoclonal antibodies that specifically react with mouse μ , γ , κ , and λ , and human μ , γ , and κ . Only hybridomas that are positive for human heavy and light chains, as well as negative for mouse chains, are identified as candidate hybridomas that express human immunoglobulin chains. Thus, candidate hybridomas are shown to have reactivity with specific antigen and to possess epitopes characteristic of a human constant region.

RNA is isolated from candidate hybridomas and used to synthesize first strand cDNA. The first strand cDNA is then ligated to a unique single-stranded oligonucleotide of predetermined sequence (oligo-X) using RNA ligase (which ligates single-stranded DNA). The ligated cDNA is then amplified in two reactions by PCR using two sets of oligonucleotide primers. Set H (heavy chain) includes an oligo that specifically anneals to either human μ or human γ 1 (depending on the results of the ELISA) and an oligo that anneals to the oligo-X sequence. This prevents bias against detection of particular V segments, including mouse V segments that may have trans-rearranged into the human minilocus. A second set of primers, Set L (light chain), includes an oligo that specifically anneals to human κ and an oligo that anneals specifically to oligo-X. The PCR products are molecularly cloned and the DNA sequence of several are determined to ascertain whether the hybridoma is producing a unique human

antibody on the basis of sequence comparison to human and murine Ig sequences.

EXAMPLE 25

This example demonstrates production of a transgenic mouse harboring a human light chain (κ) minilocus.

Human κ Minilocus transgenic mice

KC1

A 13 kb XhoI Jk2-Kk containing fragment from a phage clone (isolated from a human genomic DNA phage library by hybridization to a κ specific oligonucleotide, e.g., *supra*) was treated with Klenow enzyme and cloned into the Klenow treated HindIII site of pGP1d to produce pK-31. This destroyed the insert XhoI sites and positioned the unique polylinker derived XhoI site at the 5' end next to Jk2. A unique polylinker derived ClaI site is located between this XhoI site and the inset sequences, while a unique polylinker derived SalI site is located at the 3' end of the insert. A 7.5 kb XhoI fragment, containing Jk1 and upstream sequences, was also isolated from a human genomic DNA phage clone (isolated from a human genomic DNA phage library by hybridization to a κ specific oligonucleotide, e.g. *supra*). This 7.5 kb XhoI fragment was cloned into the SalI site of pSP72 (Promega, Madison, Wisconsin), thus destroying both XhoI sites and positioning a polylinker ClaI site 3' of Jk1. Digestion of the resulting clone with ClaI released a 4.7 kb fragment containing Jk1 and 4.5 kb of upstream sequences. This 4.7 kb fragment was cloned into the ClaI site of pK-31 to create pKcor. The remaining unique 5' XhoI site is derived from polylinker sequences. A 6.5 kb XhoI/SalI DNA fragment containing the unrearranged human VkIII gene segment 65.8 (plasmid p65.8, EXAMPLE 21) was cloned into the XhoI site of pKcor to generate the plasmid pKC1. The NotI insert of pKC1 was microinjected into 1/2 day mouse embryos to generate transgenic mice. Two independent pKC1 derived transgenic lines were established and used to breed mice containing both heavy and light chain miniloci. These lines, KC1-673 and KC1-674, were estimated by Southern blot hybridization to contain

integrations of approximately 1 and 10-20 copies of the transgenes respectively.

KC1e

The plasmid pMHE1 (EXAMPLES 13 and 18) was digested with BamHI and HindIII to excise the 2.3 kb insert containing both the mouse and human heavy chain J- μ intronic enhancers. This fragment was Klenow treated, ligated to SalI linkers (New England Biolabs, Beverly, Massachusetts), and cloned into the unique 3' SalI site of pKC1 to generate the plasmid pKC1e. The NotI insert of pKC1e was microinjected into 1/2 day mouse embryos to generate transgenic mice. Four independent pKC1e derived transgenic lines were established and used to breed mice containing both heavy and light chain miniloci. These lines, KC1e-1399, KC1e-1403, KC1e-1527, and KC1e-1536, were estimated by Southern blot hybridization to contain integrations of approximately 20-50, 5-10, 1-5, and 3-5 copies of the transgene, respectively.

pKC2

A 6.8 kb XhoI/SalI DNA fragment containing the unrearranged human VkIII gene segment 65.5 (plasmid p65.5g1, EXAMPLE 21) was cloned into the unique 5' XhoI site of pKC1 to generate the plasmid pKC2. This minilocus transgene contains two different functional VkIII gene segments. The NotI insert of pKC2 was microinjected into 1/2 day mouse embryos to generate transgenic mice. Five independent pKC2 derived transgenic lines were established and used to breed mice containing both heavy and light chain miniloci. These lines, KC2-1573, KC2-1579, KC2-1588, KC2-1608, and KC2-1610, were estimated by Southern blot hybridization to contain integrations of approximately 1-5, 10-50, 1-5, 50-100, and 5-20 copies of the transgene, respectively.

EXAMPLE 26

This example shows that transgenic mice bearing the human κ transgene can make an antigen-induced antibody

response forming antibodies comprising a functional human κ chain.

Antibody Responses Associated with Human Ig κ Light Chain

A transgenic mouse containing the HC1-57 human heavy chain and KC1e human κ transgenes was immunized with purified human soluble CD4 (a human glycoprotein antigen). Twenty μ g of purified human CD4 (NEN Research products, Westwood, MA) insolubilized by conjugation to polystyrene latex particles (Polysciences, Warrington, PA) was injected intraperitoneally in saline with dimethyldioctadecyl ammonium bromide (Calbiochem, San Diego, CA) on Day 0, followed by further injections on Day 20 and Day 34.

Retro-orbital bleeds were taken on Days 25 and 40, and screened for the presence of antibodies to CD4, containing human IgM or human IgG heavy chain by ELISA. Briefly, purified human CD4 was coated onto PVC microtitre plates at 2.5 μ g/ml and incubated with culture supernate diluted 1:4/1:5 in PBS, 0.5% Tween-20, 5% chicken serum. Plates were washed, followed by addition of horseradish peroxidase-conjugated goat antiserum specific for human IgG Fc or rabbit antiserum specific for human IgM Fc5Mu (Jackson ImmunoResearch, West Grove, PA). Presence of conjugate bound to captured antibody was determined after further washing by addition of ABTS substrate. Human μ reactive with antigen was detected in both bleeds, while there was essentially undetectable γ reactivity. The Day 40 sample was also tested for antigen-reactive human κ chain using the same assay with goat anti-human κ peroxidase conjugate (Sigma, St. Louis, MO). CD4-binding κ reactivity was detected at this time point. The assay results are shown in Fig. 48.

EXAMPLE 27

This example shows the successful generation of mice which are homozygous for functionally disrupted murine heavy and light chain loci (heavy chain and κ chain loci) and which concomitantly harbor a human heavy chain transgene and a human light chain transgene capable of productively rearranging to encode functional human heavy chains and functional human light chains. Such mice are termed "0011" mice, indicating by

the two 0's in the first two digits that the mice lack functional heavy and light chain loci and indicating by the 1's in the second two digits that the mice are hemizygous for a human heavy chain transgene and a human light chain transgene. This example shows that such 0011 mice are capable of making a specific antibody response to a predetermined antigen, and that such an antibody response can involve isotype switching.

0011/0012 Mice: Endogenous Ig Knockout + Human Ig Transgenes

Mice which were homozygous for a functionally disrupted endogenous heavy chain locus lacking a functional J_H region (designated JHD++ or JHΔ++) and also harboring the human HC1 transgene, such as the HC1-26 transgenic mouse line described *supra*, were interbred with mice homozygous for a functionally disrupted endogenous kappa chain locus lacking a functional J_H region (designated here as JKD++ or JKΔ++; see Example 9) to produce mice homozygous for functionally disrupted heavy chain and kappa chain loci (heavy chain/kappa chain knockouts), designated as JHD++/JKD++ and containing a HC1 transgene. Such mice were produced by interbreeding and selected on the basis of genotype as evaluated by Southern blot of genomic DNA. These mice, designated HC1-26+/JKD++/JHD++ mice, were interbred with mice harboring a human kappa chain transgene (lines KC2-1610, KC1e-1399, and KC1e-1527; see Example 25), and Southern blot analysis of genomic DNA was used to identify offspring mice homozygous for functionally disrupted heavy and light chain loci and also hemizygous for the HC1 transgene and the KC2 or KC1e transgene. Such mice are designated by numbers and were identified as to their genotype, with the following abbreviations: HC1-26+ indicates hemizyosity for the HC1-26 line human heavy chain minilocus transgene integration; JHD++ indicates homozygosity for J_H knockout; JKD++ indicates homozygosity for J_K knockout; KC2-1610+ indicates hemizyosity for a KC2 human κ transgene integrated as in line KC2-1610; KC1e-1527+ indicates hemizyosity for a KC1e human κ transgene integrated as in line KC1e-1527; KC1e-1399+ indicates

hemizyosity for a KC1e human κ transgene integrated as in line KC1e-1399.

The resultant individual offspring were each given a numerical designation (e.g., 6295, 6907, etc.) and each was evaluated for the presence of J_H knockout alleles, J_K knockout alleles, HC1-26 transgene, and κ transgene (KC2 or KC1e) and determined to be either hemizygous (+) or homozygous (++) at each locus. Table 10 shows the number designation, sex, and genotypes of several of the offspring mice.

Table 10

<u>ID No.</u>	<u>Sex</u>	<u>Ig Code</u>	<u>Genotype</u>
6295	M	0011	HC1-26+;JHD++;JKD++;KC2-1610+
6907	M	0011	HC1-26+;JHD++;JKD++;KC1e-1527+
7086	F	0011	HC1-26+;JHD++;JKD++;KC1e-1399+
7088	F	0011	HC1-26+;JHD++;JKD++;KC1e-1399+
7397	F	0011	HC1-26+;JHD++;JKD++;KC1e-1527+
7494	F	0012	HC1-26+;JHD++;JKD++;KC2-1610++
7497	M	0011	HC1-26+;JHD++;JKD++;KC1e-1399+
7648	F	0011	HC1-26+;JHD++;JKD++;KC2-1610+
7649	F	0012	HC1-26+;JHD++;JKD++;KC2-1610++
7654	F	0011	HC1-26+;JHD++;JKD++;KC2-1610+
7655	F	0011	HC1-26+;JHD++;JKD++;KC2-1610+
7839	F	0011	HC1-26+;JHD++;JKD++;KC1e-1399+
7656	F	0001	HC1-26-;JHD++;JKD++;KC2-1610+
7777	F	1100	Co1-2141-;JHD+;JKD+

We removed spleens from three 6 week old female mice. Mouse # 7655 was determined by Southern blot hybridization to be hemizygous for the HC1 (line 26) and KC2 (line 1610) transgene integrations, and homozygous for the JH₋ and Jk₋ targeted deletions of the mouse μ and κ J regions. Mouse #7656 was determined by Southern blot hybridization to be hemizygous for the KC2 (line 1610) transgene integration and homozygous for the JH₋ and Jk₋ targeted deletions of the mouse μ and κ J regions. Mouse # 7777 was determined by Southern blot hybridization to be hemizygous for the JH₋ and Jk₋ targeted deletions of the mouse μ and κ J regions. Because

of the recessive nature of these deletions, this mouse should be phenotypically wild-type.

Expression of Endogenous Ig Chains in 0011 Mice

FACS analysis using a panel of antibodies reactive with either human μ , mouse μ , human κ , mouse κ , or mouse λ was used to sort lymphocytes explanted from (1) a wildtype mouse (7777), (2) a 0001 mouse homozygous for heavy chain and kappa knockout alleles and harboring a human light chain transgene (7656), and (3) a 0011 mouse homozygous for heavy chain and kappa knockout alleles and harboring a human light chain transgene and a human heavy chain transgene (7655).

We prepared single cell suspensions from spleen and lysed the red cells with NH_4Cl , as described by Mishell and Shiigi (Mishell, B.B. & Shiigi, S.M. (eds) Selected Methods in Cellular Immunology. W.H. Freeman & Co., New York, 1980).

The lymphocytes are stained with the following reagents:

propidium iodide (Molecular Probes, Eugene, OR), FITC conjugated anti-human IgM (clone G20-127; Pharmingen, San Diego, CA), FITC conjugated anti-mouse IgM (clone R6-60.2; Pharmingen, San Diego, CA), phycoerythrin conjugated anti-human Igk (clone HP6062; CalTag, South San Francisco, CA), FITC conjugated anti-mouse Ig λ (clone R26-46; Pharmingen, San Diego, CA) FITC conjugated anti-mouse B220 (clone RA3-6B2; Pharmingen, San Diego, CA), and Cy-Chrome conjugated anti-mouse B220 (clone RA3-6B2; Pharmingen, San Diego, CA). We analyzed the stained cells using a FACScan flow cytometer and LYSIS II software (Becton Dickinson, San Jose, CA).

Macrophages and residual red cells are excluded by gating on forward and side scatter. Dead cells are excluded by gating out propidium iodide positive cells. The flow cytometric data in Figs. 49 and 50 confirms the Southern blot hybridization data and demonstrates that mouse #7655 expresses both human μ and human κ and relatively little if any mouse μ or mouse κ . Nevertheless a significant fraction of the B cells (about 70-80%) appear to express hybrid Ig receptors consisting of human heavy and mouse λ light chains.

Fig. 49 shows the relative distribution of B cells expressing human μ or mouse μ on the cell surface; 0011 mouse (7655) lymphocytes are positive for human μ but relatively lack mouse μ ; 0001 mouse (7656) lymphocytes do not express much human μ or mouse μ ; wildtype mouse (7777) lymphocytes express mouse μ but lack human μ .

Fig. 50 shows the relative distribution of B cells expressing human κ or mouse κ on the cell surface; 0011 mouse (7655) lymphocytes are positive for human κ but relatively lack mouse κ ; 0001 mouse (7656) lymphocytes do not express much human κ or mouse κ ; wildtype mouse (7777) lymphocytes express mouse κ but lack human κ .

Fig. 51 shows the relative distribution of B cells expressing mouse λ on the cell surface; 0011 mouse (7655) lymphocytes are positive for mouse λ ; 0001 mouse (7656) lymphocytes do not express significant mouse λ ; wildtype mouse (7777) lymphocytes express mouse λ but at a relatively lower level than the 0011 mouse (7655).

Fig. 52 shows the relative distribution of B cells positive for endogenous mouse λ as compared to human κ (transgene-encoded). The upper left panel shows the results of cells from a wildtype mouse possessing functional endogenous heavy and light chain alleles and lacking human transgene(s); the cells are positive for mouse lambda. The upper right panel shows cells from a mouse (#5822) having a κ knockout background (JKD++) and harboring the human κ transgene intergration of the KC1e-1399 line; the cells are positive for human κ or mouse λ in roughly proportional amounts. The lower left panel shows cells from a mouse (#7132) having a κ knockout background (JKD++) and harboring the human κ transgene intergration of the KC2-1610 line; more cells are positive for mouse λ than for human κ , possibly indicating that the KC2-1610 transgene integration is less efficient than the KC1e-1399 transgene integration. The lower right panel shows cells from a mouse harboring a human κ minilocus transgene (KCo4) and lacking a functional endogenous murine κ allele. The data presented in Fig. 52 also demonstrates the variability of phenotypic expression between

transgenes. Such variability indicates the desirability of selecting for individual transgenes and/or transgenic lines which express one or more desired phenotypic features resulting from the integrated transgene (e.g., isotype switching, high level expression, low murine Ig background). Generally, single or multiple transgene species (e.g., pKC1e, pKC2, KCo4) are employed separately to form multiple individual transgenic lines differing by: (1) transgene, (2) site(s) of transgene integration, and/or (3) genetic background. Individual transgenic lines are examined for desired parameters, such as: (1) capability to mount an immune response to a predetermined antigen, (2) frequency of isotype switching within transgene-encoded constant regions and/or frequency of trans-switching to endogenous (e.g., murine) Ig constant region genes, (3) expression level of transgene-encoded immunoglobulin chains and antibodies, (4) expression level of endogenous (e.g., murine) immunoglobulin immunoglobulin sequences, and (5) frequency of productive VDJ and VJ rearrangement. Typically, the transgenic lines which produce the largest concentrations of transgene-encoded (e.g., human) immunoglobulin chains are selected; preferably, the selected lines produce about at least 40 µg/ml of transgene-encoded heavy chain (e.g., human µ or human γ) in the serum of the transgenic animal and/or about at least 100 µg/ml of transgene-encoded light chain (e.g., human κ).

Mice were examined for their expression of human and murine immunoglobulin chains in their unimmunized serum and in their serum following immunization with a specific antigen, human CD4. Fig. 53 shows the relative expression of human µ, human γ, murine µ, murine γ, human κ, murine κ, and murine λ chains present in the serum of four separate unimmunized 0011 mice of various genotypes (nt = not tested); human κ predominates as the most abundant light chain, and human µ and murine γ (putatively a product of trans-switching) are the most abundant heavy chains, with variability between lines present, indicating the utility of a selection step to identify advantageous genotypic combinations that minimize expression of murine chains while allowing expression of human

chains. Mice #6907 and 7088 show isotype switching (cis-switching within the transgene) from human μ to human γ .

Fig. 54 shows serum immunoglobulin chain levels for human μ (hup), human γ (huy), human κ (huk), murine μ (msp), murine γ (msy), murine κ (msk), and murine λ (msl) in mice of the various 0011 genotypes.

Specific Antibody Response in 0011 Mice

An 0011 mouse (#6295) was immunized with an immunogenic dose of human CD4 according to the following immunization schedule: Day 0, intraperitoneal injection of 100 μ l of CD4 mouse immune serum; Day 1, inject 20 μ g of human CD4 (American Bio-Tech) on latex beads with DDA in 100 μ l; Day 15 inject 20 μ g of human CD4 (American Bio-Tech) on latex beads with DDA in 100 μ l; Day 29 inject 20 μ g of human CD4 (American Bio-Tech) on latex beads with DDA in 100 μ l; Day 43 inject 20 μ g of human CD4 (American Bio-Tech) on latex beads with DDA in 100 μ l.

Fig. 55 shows the relative antibody response to CD4 immunization at 3 weeks and 7 weeks demonstrating the presence of human μ , human κ , and human γ chains in the anti-CD4 response. Human γ chains are present at significantly increased abundance in the 7 week serum, indicating that cis-switching within the heavy chain transgene (isotype switching) is occurring in a temporal relationship similar to that of isotype switching in a wildtype animal.

Fig. 56 shows a schematic compilation of various human heavy chain and light chain transgenes.

EXAMPLE 28

This example provides for the targeted knockout of the murine λ light chain locus.

Targeted Inactivation of the Murine Lambda Light Chain Locus

Unlike the Ig heavy and kappa light chain loci, the murine $V\lambda J\lambda$ and $C\lambda$ gene segments are not grouped into 3 families arranged in a 5' to 3' array, but instead are interspersed. The most 5' portion consists of two V segments ($V\lambda 2$ and $V\lambda X$) which are followed, proceeding in a 3'

direction, by two constant region exons, each associated with its own J segment (J λ 2C λ 2 and the pseudogene J λ 4C λ 4). Next is the most extensively used V segment (V λ 1) which is followed by the second cluster of constant region exons (J λ 3C λ 3 and J λ 1C λ 1,). Overall the locus spans approximate 200 kb, with intervals of ~20-90 kb between the two clusters.

Expression of the lambda locus involves rearrangement of V λ 2 or V λ X predominantly to J λ 2 and only rarely further 3' to J λ 3 or J λ 1. V λ 1 can recombine with both J λ 3 and J λ 1. Thus the lambda locus can be mutated in order to fully eliminate recombination and expression of the locus.

The distance between the two lambda gene clusters makes it difficult to inactivate expression of the locus via the generation of a single compact targeted deletion, as was used in inactivating the murine Ig heavy and kappa light chain loci. Instead, a small single deletion which would eliminate expression lambda light chains spans approximately 120 kb, extending from J λ 2C λ 2 to J λ 1C λ 1 (Fig. 57). This removes all of the lambda constant region exons as well as the V λ 1 gene segment, ensuring inactivation of the locus.

Replacement type targeting vectors (Thomas and Capecchi (1987) op.cit) are constructed in which the deleted 120 kb is replaced with the selectable marker gene, neo, in a PGK expression cassette. The marker is embedded within genomic lambda sequences flanking the deletion to provide homology to the lambda locus and can also contain the HSV-*tk* gene, at the end of one of the regions of homology, to allow for enrichment for cells which have homologously integrated the vectors. Lambda locus genomic clone sequences are obtained by screening of a strain 129/Sv genomic phage library isogenic to the ES line being targeted, since the use of targeting vectors isogenic to the chromosomal DNA being targeted has been reported to enhance the efficiency of homologous recombination. Targeting vectors are constructed which differ in their lengths of homology to the lambda locus. The first vector (vector 1 in Fig. 58) contains the marker gene flanked by total of approximately 8-12 kb of lambda locus sequences. For targeting events in which replacement vectors

mediate addition or detection of a few kb of DNA this has been demonstrated to be a more than sufficient extent of homology (Hasty et al. (1991) op.cit; Thomas et al. (1992) op.cit.). Vectors with an additional approximately 40-60 kb of flanking lambda sequence are also constructed (vector 2 in Fig. 58). Human Ig miniloci of at least 80 kb are routinely cloned and propagated in the plasmid vector pGP1 (Taylor et al. (1993) op.cit).

An alternative approach for inactivation of the lambda locus employs two independent mutations, for example mutations of the two constant region clusters or of the two V region loci, in the same ES cell. Since both constant regions are each contained within ~6 kb of DNA, whereas one of the V loci spans ~19 kb, targeting vectors are constructed to independently delete the J λ 2C λ 2/J λ 4C λ 4 and the J λ 3C λ 3/J λ 1C λ 1 loci. As shown in Fig. 58, each vector consists of a selectable marker (e.g., neo or pac) in a PGK expression cassette, surrounded by a total of ~8-12 kb of lambda locus genomic DNA blanking each deletion. The HSV-tk gene can be added to the targeting vectors to enrich for homologous recombination events by positive-negative selection. ES cells are targeted sequentially with the two vectors, such that clones are generated which carry a deletion of one of the constant region loci; these clones are then targeted sequentially with the two vectors, such that clones will be generated which carry a deletion of one of the constant region loci, and these clones are then targeted to generate a deletion of the remaining functional constant region cluster. Since both targeting events are thus being directed to the same cell, it is preferable to use a different selectable marker for the two targetings. In the schematic example shown in Fig. 58, one of the vectors contains the neo gene and the other the pac (puromycin N-acetyl transferase) gene. A third potential dominant selectable marker is the hyg (hygromycin phosphotransferase) gene. Both the pac and hyg genes can be inserted into the PGK expression construct successfully used for targeting the neo gene into the Ig heavy and kappa light chain loci. Since the two lambda constant region

clusters are tightly linked, it is important that the two mutations reside on the same chromosome. There preferably is a 50% probability of mutating the same allele by two independent targeting events, and linkage of the mutations is established by their co-segregation during breeding of chimeras derived from the doubly targeted ES cells.

EXAMPLE 29

This example provides for the targeted knockout of the murine heavy chain locus.

Targeted Inactivation of the Murine Heavy Chain Locus

A homologous recombination gene targeting transgene having the structure shown in Fig. 59 is used to delete at least one and preferably substantially all of the murine heavy chain locus constant region genes by gene targeting in ES cells. Fig. 59 shows a general schematic diagram of a targeting transgene. Segment (a) is a cloned genomic DNA sequence located upstream of the constant region gene(s) to be deleted (i.e., proximal to the J_H genes); segment (b) comprises a positive selection marker, such as pgk-neo; segment (c) is a cloned genomic DNA sequence located downstream of the constant region gene(s) to be deleted (i.e., distal to the constant region gene(s) and J_H genes); and segment (d), which is optional, comprises a negative selection marker gene (e.g., HSV-tk). Fig. 60 shows a map of the murine heavy chain locus as taken from Immunoglobulin Genes, Honjo, T, Alt, FW, and Rabbits TH (eds.) Academic Press, NY (1989) p. 129.

A targeting transgene having a structure according to Fig. 59, wherein: (1) the (a) segment is the 11.5 kb insert of clone JH8.1 (Chen et al. (1993) Int. Immunol. 5: 647) or an equivalent portion comprising about at least 1-4 kb of sequence located upstream of the murine C_H gene, (2) the (b) segment is pgk-neo as described *supra*, (3) the (c) segment comprises the 1674 bp sequence shown in Fig. 61 or a 4-6 kb insert isolated from a phage clone of the mouse $C\alpha$ gene isolated by screening a mouse genomic clone library with the end-labeled oligonucleotide having the sequence:

5'-gtg ttg cgt gta tca gct gaa acc tgg aaa cag ggt gac cag-3'
and (4) the (d) segment comprises the HSV-tk expression cassette described *supra*.

Alternatively, a stepwise deletion of one or more heavy chain constant region genes is performed wherein a first targeting transgene comprises homology regions, i.e., segments (a) and (c), homologous to sequences flanking a constant region gene or genes, a first species of positive selection marker gene (pgk-neo), and an HSV-tk negative selection marker. Thus, the (a) segment can comprise a sequence of at least about 1-4 kb and homologous to a region located upstream of Cy3 and the (c) segment can comprise a sequence of at least about 1-4 kb and homologous to a region located upstream of Cy2a. This targeting transgene deletes the Cy3, Cy1, Cy2b, and Cy2a genes. This first targeting transgene is introduced into ES cells and correctly targeted recombinants are selected (e.g., with G418), producing a correctly targeted C region deletion. Negative selection for loss of the HSV-tk cassette is then performed (e.g., with ganciclovir or FIAU). The resultant correctly targeted first round C deletion recombinants have a heavy chain locus lacking the Cy3, Cy1, Cy2b, and Cy2a genes.

A second targeting transgene comprises homology regions, i.e., segments (a) and (c), homologous to sequences flanking a constant region gene or genes, a second species of positive selection marker gene different than the first species (e.g., gpt or pac), and an HSV-tk negative selection marker. Thus, the (a) segment can comprise a sequence of at least about 1-4 kb and homologous to a region located upstream of C ϵ and the (c) segment can comprise a sequence of at least about 1-4 kb and homologous to a region located upstream of C α . This targeting transgene deletes the C ϵ and C α genes.

This second targeting transgene is introduced into the correctly targeted C-region recombinant ES cells obtained from the first targeting event. Cells which are correctly targeted for the second knockout event (i.e., by homologous recombination with the second targeting transgene) are selected for with a selection drug that is specific for the

second species of positive selection marker gene (e.g., mycophenolic acid to select for gpt; puromycin to select for pac). Negative selection for loss of the HSV-tk cassette is then performed (e.g., with ganciclovir or FIAU). These resultant correctly targeted second round C region recombinants have a heavy chain locus lacking the Cy3, Cy1, Cy2b, Cy2a, C ϵ , and C α genes.

Correctly targeted first-round or second-round recombinant ES cells lacking one or more C region genes are used for blastocyst injections as described (*supra*) and chimeric mice are produced. Germline transmission of the targeted heavy chain alleles is established, and breeding of the resultant founder mice is performed to generate mice homozygous for C-region knockouts. Such C-region knockout mice have several advantages as compared to J_H knockout mice; for one example, C-region knockout mice have diminished ability (or completely lack the ability) to undergo trans-switching between a human heavy chain transgene and an endogenous heavy chain locus constant region, thus reducing the frequency of chimeric human/mouse heavy chains in the transgenic mouse. Knockout of the murine gamma genes is preferred, although μ and delta are frequently also deleted by homologous targeting. C-region knockout can be done in conjunction with other targeted lesions in the endogenous murine heavy chain locus; a C-region deletion can be combined with a J_H knockout to preclude productive VDJ rearrangement of the murine heavy chain locus and to preclude or reduce trans-switching between a human heavy chain transgene and the murine heavy chain locus, among others. For some embodiments, it may be desirable to produce mice which specifically lack one or more C-region genes of the endogenous heavy chain locus, but which retain certain other C-region genes; for example, it may be preferable to retain the murine C α gene to allow to production of chimeric human/mouse IgA by trans-switching, if such IgA confers an advantageous phenotype and does not substantially interfere with the desired utility of the mice.

This example demonstrates ex vivo depletion of lymphocytes expressing an endogenous (murine) immunoglobulin from a lymphocyte sample obtained from a transgenic mouse harboring a human transgene. The lymphocytes expressing murine Ig are selectively depleted by specific binding to an anti-murine immunoglobulin antibody that lacks substantial binding to human immunoglobulins encoded by the transgene(s).

Ex Vivo Depletion of Murine Ig-Expressing B-cells

A mouse homozygous for a human heavy chain minilocus transgene (HC2) and a human light chain minilocus transgene (KCo4) is bred with a C57BL/6 (B6) inbred mouse to obtain 2211 mice (i.e., mice which: are homozygous for a functional endogenous murine heavy chain locus, are homozygous for a functional endogenous murine light chain locus, and which possess one copy of a human heavy chain transgene and one copy of a human light chain transgene). Such 2211 mice also express B6 major and minor histocompatibility antigens. These mice are primed with an immunogenic dose of an antigen, and after approximately one week spleen cells are isolated. B cells positive for murine Ig are removed by solid phase-coupled antibody-dependent cell separation according to standard methods (Wysocki et al. (1978) Proc. Natl. Acad. Sci. (U.S.A.) 75: 2844; MACS magnetic cell sorting, Miltenyi Biotec Inc., Sunnyvale, CA), followed by antibody-dependent complement-mediated cell lysis (Selected Methods in Cellular Immunology, Mishell BB and Shiigi SM (eds.), W.H. Freeman and Company, New York, 1980, pp.211-212) to substantially remove residual cells positive for murine Ig. The remaining cells in the depleted sample (e.g., T cells, B cells positive for human Ig) are injected i.v., preferably together with additional anti-murine Ig antibody to deplete arising B cells, into a SCID/B6 or RAG/B6 mouse. The reconstituted mouse is then further immunized for the antigen to obtain antibody and affinity matured cells for producing hybridoma clones.

EXAMPLE 31

Production of Fully Human Antibodies in Somatic Chimeras

A method is described for producing fully human antibodies in somatic chimeric mice. These mice are generated by introduction of embryonic stem (ES) cells, carrying human immunoglobulin (Ig) heavy and light chain transgenes and lacking functional murine Ig heavy and kappa light chain genes, into blastocysts from RAG-1 or RAG-2 deficient mice.

RAG-1 and RAG-2 deficient mice (Mombaerts et al. (1992) Cell 68: 869; Shinkai et al. (1992) Cell 68: 855) lack murine B and T cells due to an inability to initiate VDJ rearrangement and to assemble the gene segments encoding Igs and T cell receptors (TCR). This defect in B and T cell production can be complemented by injection of wild-type ES cells into blastocysts derived from RAG-2 deficient animals. The resulting chimeric mice produce mature B and T cells derived entirely from the injected ES cells (Chen et al. (1993) Proc. Natl. Acad. Sci. USA 90: 4528).

Genetic manipulation of the injected ES cells is used for introducing defined mutations and/or exogenous DNA constructs into all of the B and/or T cells of the chimeras. Chen et al. (1993), Proc. Natl. Acad. Sci. USA 90:4528-4532) generated ES cells carrying a homozygous inactivation of the Ig heavy chain locus, which, when injected into RAG blastocysts, produced chimeras which made T cells in the absence of B cells. Transfection of a rearranged murine heavy chain into the mutant ES cells results in the rescue of B cell development and the production of both B and T cells in the chimeras.

Chimeric mice which express fully human antibodies in the absence of murine Ig heavy chain or kappa light chain synthesis can be generated. Human Ig heavy and light chain constructs are introduced into ES cells homozygous for inactivation of both the murine Ig heavy and kappa light chain genes. The ES cells are then injected into blastocysts derived from RAG2 deficient mice. The resulting chimeras contain B cells derived exclusively from the injected ES cells which are incapable of expressing murine Ig heavy and kappa light chain genes but do express human Ig genes.

Generation of ES cells Homozygous for Inactivation of the Immunoglobulin Heavy and Kappa Light Chain Genes

Mice bearing inactivated Ig heavy and kappa light chain loci were generated by targeted deletion, in ES cells, of Ig J_H and J_K/C_K sequences, respectively according to known procedures (Chen et al. (1993) EMBO J. 12: 821; and Chen et al. (1993) Int. Immunol. op.cit). The two mutant strains of mice were bred together to generate a strain homozygous for inactivation of both Ig loci. This double mutant strain was used for derivation of ES cells. The protocol used was essentially that described by Robertson (1987, in Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, p. 71-112, edited by E.J. Robertson, IRL Press). Briefly, blastocysts were generated by natural matings of homozygous double mutant mice. Pregnant females were ovariectomized on day 2.5 of gestation and the "delayed" blastocysts were flushed from the uterus on day 7 of gestation and cultured on feeder cells, to help maintain their undifferentiated state. Stem cells from the inner cell mass of the blastocysts, identifiable by their morphology, were picked, dissociated, and passaged on feeder cells. Cells with a normal karyotype were identified, and male cell lines will be tested for their ability to generate chimeras and contribute to the germ cells of the mouse. Male ES cells are preferable to female lines since a male chimera can produce significantly more offspring.

Introduction of Human Ig Genes into Mouse Ig Heavy and Kappa Light Chain Deficient ES cells

Human immunoglobulin heavy and light chain genes are introduced into the mutant ES cells as either minilocus constructs, such as HC2 and KC-C04, or as YAC clones, such as J1.3P. Transfection of ES cells with human Ig DNAs is carried out by techniques such as electroporation or lipofection with a cationic lipid. In order to allow for selection of ES cells which have incorporated the human DNA, a selectable marker either is ligated to the constructs or is co-transfected with the constructs into ES cells. Since the mutant ES cells contain the neomycin phosphotransferase (neo) gene as a result

of the gene targeting events which generated the Ig gene inactivations, different selectable markers, such as hygromycin phosphotransferase (hyg) or puromycin N-acetyl transferase (pac), are used to introduce the human Ig genes into the ES cells.

The human Ig heavy and light chain genes can be introduced simultaneously or sequentially, using different selectable markers, into the mutant ES cells. Following transfection, cells are selected with the appropriate selectable marker and drug-resistant colonies are expanded for freezing and for DNA analysis to verify and analyze the integration of the human gene sequences.

Generation of Chimeras

ES clones containing human Ig heavy and light chain genes are injected into RAG-2 blastocysts as described (Bradley, A. (1987), in Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, p. 113-151, edited by E.J. Robertson, IRL Press) and transferred into the uteri of pseudopregnant females. Offspring are screened for the presence of human antibodies by ELISA assay of serum samples. Positive animals are used for immunization and the production of human monoclonal antibodies.

EXAMPLE 32

This example describes the introduction, via homologous recombination in ES cells, of a targeted frameshift mutation into the murine heavy chain locus leading to a deletion of B cells which undergo switch recombination. The frameshifted mice are suitable hosts for harboring non-murine (e.g., human) transgenes encoding human sequence immunoglobulins.

The novel frameshifted mice can be used for expressing non-murine (e.g., human) sequence immunoglobulins encoded by heavy chain transgene(s) and/or light chain transgene(s), and for the isolation of hybridomas expressing class-switched, affinity matured, human sequence antibodies from introduced transgenes, among other uses. A frameshift is introduced into one of the four mouse JH gene segments and

into the first exon of the mouse μ gene. The two introduced frameshift mutations compensate for each other thus allowing for the expression of fully functional murine μ heavy chain when a B cell uses the frameshifted JH for a functional VDJ joint. None of the other three JH segments can be used for functional VDJ joining because of the frameshift in μ , which is not compensated in the remaining JH genes. Alternatively, compensating frameshifts can be engineered into multiple murine JH genes.

A mouse homozygous for a compensated, frameshifted immunoglobulin heavy chain allele has an approximately physiological level of peripheral B cells, and an approximately physiological level of serum IgM comprising both murine and human μ . However, B cells recruited into germinal centers frequently undergo a class switch to a non- μ isotype. Such a class switch in B cells expressing the endogenous murine μ chain leads to the expression of a non-compensated frameshift mRNA, since the remaining non- μ C_H genes do not possess a compensating frameshift. The resulting B cells do not express a B cell receptor and are deleted. Hence, B cells expressing a murine heavy chain are deleted once they reach the stage of differentiation where isotype switching occurs. However, B cells expressing heavy chains encoded by a non-murine (e.g., human) transgene capable of isotype switching and which does not contain such isotype-restrictive frameshifts are capable of further development, including isotype switching and/or affinity maturation, and the like.

Therefore, the frameshifted mouse has an impaired secondary response with regard to murine heavy chain (μ) but a significant secondary response with regard to transgene-encoded heavy chains. If a heavy chain transgene that is capable of undergoing class switching is introduced into this mutant background, the non-IgM secondary response is dominated by transgene expressing B cells. It is thus possible to isolate affinity matured human sequence immunoglobulin expressing hybridomas from these frameshifted mice. Moreover, the frameshifted mice generally possess immunoprotective levels of murine IgM, which may be advantageous where the

human heavy chain transgene can encode only a limited repertoire of variable regions.

For making hybridomas secreting human sequence monoclonal antibodies, transgenic mutant mice are immunized; their spleens fused with a myeloma cell line; and the resulting hybridomas screened for expression of the transgene encoded human non- μ isotype. Further, the frameshifted mouse may be advantageous over a JH deleted mouse because it will contain a functional μ switch sequence adjacent to a transcribed VDJ which serves as an active substrate for cis-switching (Gu et al. (1993) Cell 73: 1155); thus reducing the level of trans-switched B cells that express chimeric human/mouse antibodies.

Construction of Frameshift Vectors

Two separate frameshift vectors are built. One of the vectors is used to introduce 2 nucleotides at the 3' end of the mouse J4 gene segment, and one of the vectors is used to delete those same two nucleotides from the 5' end of exon 1 of the mouse μ gene.

1. JH vector.

A 3.4 kb XhoI/EcoRI fragment covering the mouse heavy chain J region and the μ intronic enhancer is subcloned into a plasmid vector that contains a neomycin resistance gene as well as a herpes thymidine kinase gene under the control of a phosphoglycerate kinase promoter (tk/neo cassette; Hasty et al., (1991) Nature 350: 243). This clone is then used as a substrate for generating 2 different PCR fragments using the following oligonucleotide primers:

o-A1 5'- cca cac tct gca tgc tgc aga agc ttt tct gta -
3' (SEQ ID NO:161)

o-A2 5'- ggt gac tga ggt acc ttg acc cca gta gtc cag -
3' (SEQ ID NO:162)

o-A3 5'- ggt tac ctc agt cac cgt ctc ctc aga ggt aag aat
ggc ctc -3' (SEQ ID NO:163)

o-A4 5'- agg ctc cac cag acc tct cta gac agc aac tac -
3' (SEQ ID NO:164)

Oligonucleotides o-A1 and o-A2 are used to amplify a 1.2 kb fragment which is digested with SphI and KpnI. Oligonucleotides o-A3 and o-A4 are used to amplify a 0.6 kb fragment which is digested with KpnI and XbaI. These two digested fragments are then cloned into SphI/XbaI digested plasmid A to produce plasmid B.

Plasmid B contains the 2 nucleotide insertion at the end of the J4 and, in addition, contains a new KpnI site upstream of the insertion. The KpnI site is used as a diagnostic marker for the insertion.

Additional flanking sequences may be cloned into the 5' XhoI site and the 3' EcoRI site of plasmid B to increase its homologous recombination efficiency. The resulting plasmid is then digested with SphI, or another restriction enzyme with a single site within the insert, and electroporated into embryonic stem cells which are then selected with G418 as described by Hasty et al. (1991) op.cit. Homologous recombinants are identified by Southern blot hybridization and then selected with FIAU as described by Hasty et al. to obtain deleted subclones which contain only the 2 base pair insertion and the new KpnI site in JH4. These are identified by Southern blot hybridization of KpnI digested DNA and confirmed by DNA sequence analysis of PCR amplified JH4 DNA.

The resulting mouse contains a JH4 segment that has been converted from the unmutated sequence:

...TGGGGTCAAGGAACCTCAGTCACCGTCTCCTCAG__gtaagaatggcctctcc...

TrpGlyGlnGlyThrSerValThrValSerSerGlu

(SEQ ID NOS:165 and 166, respectively)

to the mutant sequence:

...TGGGGTCAAGGTACCTCAGTCACCGTCTCCTCAGAGgtaagaatggcctctcc...

TrpGlyGlnGlyThrSerValThrValSerSerGlu

(SEQ ID NOS:167 and 168, respectively)

μ Exon 1 Vector

2000

Using similar in vitro mutagenesis methodology described above to engineer a two base pair insertion into the JH4 gene segment, PCR products and genomic subclones are assembled to create a vector containing a two base pair deletion at the 5' end of the first μ exon. In addition, to mark the mutation, a new XmnI site is also introduced downstream by changing an A to a G.

The sequence of the unmutated μ gene is:

...ctggtcctcagAGAGTCAGTCCTTCCCAATGTCTTCCCCCTCGTC...

GluSerGlnSerPheProAsnValPheProLeuVal

(SEQ ID NOS:169 and 170, respectively).

The sequence of the mutated μ gene is:

XmnI

...ctggtcctcag__AGTCAGTCCTTCCCGAATGTCTTCCCCCTCGTC...

SerGlnSerPheProAsnValPheProLeuVal

(SEQ ID NOS:171 and 172, respectively).

The homologous recombination vector containing the mutant sequence is linearized and electroporated into an ES cell line containing the JH4 insertion. Homologous recombinants are identified from neomycin-resistant clones. Those homologous recombinants that contain the frameshift insertion on the same chromosome as the JH4 insertion are identified by Southern blot hybridization of KpnI/BamHI digested DNA. The JH4 insertion is associated with a new KpnI site that reduces the size of the J- μ intron containing KpnI/BamHI fragment from the wild type 11.3 kb to a mutant 9 kb. The resulting clones are then selected for deletion of the inserted tk/neo cassette using FIAU. Clones containing the mutant μ exon are identified by Southern blot hybridization of XmnI digested DNA. The mutation is confirmed by DNA sequence analysis of PCR amplified μ exon1 DNA.

Generation of Frameshifted Mice

The ES cell line containing both the two base pair insertion in JH4, and the two base pair deletion in μ exon 1, is then introduced into blastocyst stage embryos which are inserted into pseudopregnant females to generate chimeras. Chimeric animals are bred to obtain germline transmission, and

the resulting animals are bred to homozygosity to obtain mutant animals homozygous for compensated frameshifted heavy chain loci and having impaired secondary humoral immune responses in B cells expressing murine heavy chains.

A human heavy chain transgene, such as for example pHCl or pHCl2 and the like, may be bred into the murine heavy chain frameshift background by crossbreeding mice harboring such a human transgene into mice having the frameshifted murine IgH locus. Via interbreeding and backcrossing, mice homozygous at the murine IgH locus for μ -compensated frameshifted murine IgH alleles (i.e., capable of compensated in-frame expression of only murine μ and not murine non- μ chains) and harboring at least one integrated copy of a functional human heavy chain transgene (e.g., pHCl or pHCl2) are produced. Such mice may optionally contain knockout of endogenous murine κ and/or λ loci as described supra, and may optionally comprise a human or other non-murine light chain transgene (e.g., pKCl1e, pKCl2, and the like).

Alternatively, the human transgene(s) (heavy and/or light) may comprise compensating frameshifts, so that the transgene J gene(s) contain a frameshift that is compensated by a frameshift in the transgene constant region gene(s). Trans-switching to the endogenous constant region genes is uncompensated and produces a truncated or nonsense product; B cells expressing such uncompensated trans-switched immunoglobulins are selected against and depleted.

EXAMPLE 33

Endogenous Heavy Chain Inactivation by D Region Ablation

This example describes a positive-negative selection homologous recombination vector for replacing the mouse germline immunoglobulin heavy chain D region with a nonfunctional rearranged VDJ segment. The resulting allele functions within a B cell as a normal non-productive allele, with the allele undergoing intra-allele heavy chain class switching, thereby reducing the level of trans-switching to an active transgene locus.

D Region Targeting Construct

An 8-15 kb DNA fragment located upstream of the murine D region is isolated and subcloned from a mouse strain 129 phage library using an oligonucleotide probe comprising approximately 50 consecutive nucleotides of the published sequence for the DFL16.1 segment listed in GenBank. DFL16.1 is the upstream D segment (i.e., proximal to the V region gene cluster and distal to the constant region gene cluster).

Similarly, a 9.5 kb BamHI fragment containing JH3, JH4, E μ , S μ , and the first two coding exons of the μ constant region is isolated and subcloned from a mouse strain 129 genomic phage library.

A 5-10 kb rearranged VDJ is then isolated from a mouse hybridoma (any strain) and a synthetic linker containing a stop codon is inserted into the J segment. The stop linker within the J is preferable to an out-of-frame VDJ junction because of the possibility of V replacement rearrangements.

These three fragments are assembled together with a PGKneo positive selection cassette and a PGKHSVtk negative selection cassette to form a positive-negative selection vector for eliminating the mouse D region in 129-derived ES cells (e.g., AB1) by homologous recombination. The targeting vector is formed by ligating the 8-15 kb DNA fragment to the positive selection cassette (e.g., PGKneo), which is itself ligated to the rearranged 5-10 kb rearranged VDJ, which is itself ligated to the 9.5 kb BamHI fragment; the negative selection cassette (e.g., PGKHSVtk) is then ligated at either end of the targeting construct. The construction of such a D region targeting vector is shown schematically in Fig. 63.

The D region targeting construct is transferred into AB1 ES cells, positive and negative selection is performed as described above, and correctly targeted ES cells are cloned. The correctly targeted ES cell clones are used for blastocyst injections and chimeric mice are produced. The chimeric mice are bred to produce founder mice harboring a D-region inactivated heavy chain allele. Interbreeding of offspring is performed to produce homozygotes lacking a functional endogenous heavy chain locus. Such homozygotes are used to crossbreed to mice harboring human Ig transgenes (e.g., pHCl,

pHC2, pKC2, pKC1e, KCo4) to yield (by further backcrossing to the homozygotes lacking a functional D-region) mice lacking a functional endogenous heavy chain locus and harboring a human heavy transgene (and preferably also a human light chain transgene). In embodiments where some functional endogenous light chain loci remain (e.g., λ loci), it is generally preferred that transgenes contain transcriptional control sequences that direct high level expression of human light chain (e.g., κ) polypeptides, and thus allow the transgene locus to compete effectively with the remaining endogenous light chain (e.g., λ) loci. For example, the Co4 kappa light chain transgene is generally preferred as compared to pKC1 with regard to the ability to compete effectively with the endogenous λ loci in the transgenic animal.

EXAMPLE 34

This example describes expansion of the human light chain transgene V gene repertoire by co-injection of a human κ light chain minilocus and a yeast artificial chromosome comprising a portion of the human V_k locus.

Introduction of Functional Human Light Chain V Segments by Co-Injection of V_k -Containing YAC DNA and a κ Minilocus

An approximately 450 kb YAC clone containing part of the human V_k locus was obtained as a non-amplified YAC DNA from clone 4x17E1 of the publicly available ICRF YAC library (Larin et al. (1991) Proc. Natl. Acad. Sci. (U.S.A.) 88: 4123; Genome Analysis Laboratory, Imperial Cancer Research Fund, London, UK). The 450 kb YAC clone was isolated without prior amplification by standard pulsed-field gel electrophoresis as per the manufacturer's specifications (CHEF DR-II electrophoresis cell, Bio-Rad Laboratories, Richmond, CA). Six individual pulse field gels were stained with ethidium bromide and the gel material containing the YAC clone DNA was excised from the gel and then embedded in a new (low melting point agarose in standard gel buffer) gel cast in a triangular gel tray. The resulting triangular gel (containing the six excised YAC-containing gel blocks) was extended at the apex with a narrow agarose gel with 2 M NaOAc in addition to the

standard electrophoresis buffer. The gel was then placed in an electrophoresis chamber immersed in standard gel buffer. The Y-shaped gel former rises above the surface of the buffer so that current can only flow to the narrow high salt gel portion. A plexiglas block was placed over the high salt gel slice to prevent diffusion of the NaOAc into the buffer. The YAC DNA was then electrophoresed out of the original excised gel sliced (embedded) and into the narrow high salt gel portion. At the point of transition from the low salt gel to the high salt gel, there is a resistance drop that effectively halts the migration of the DNA at the apex of the triangular gel.

Following electrophoresis and staining with ethidium bromide, the concentrated YAC DNA was cut away from the rest of the gel and the agarose was digested with GELase (EpiCentre Technologies, Madison, Wisconsin). Cesium chloride was then added to the resultant YAC-containing liquid to obtain a density of 1.68 g/ml. This solution was centrifuged at 37,000 rpm for 36 hours to separate the YAC DNA from any contaminating material. 0.5 ml fractions of the resulting density gradient were isolated and the peak DNA fraction was dialyzed against 5 mM Tris (pH 7.4), 5 mM NaCl, 0.1 M EDTA. Following dialysis, the concentration of the resulting 0.65 ml solution of YAC DNA was found to contain 2 µg/ml of DNA. This YAC DNA was mixed with purified DNA insert from plasmids pKC1B and pKV4 at a ratio of 20:1:1 (micrograms YAC4x17E1:KC1B:KV4). The resulting 2 µg/ml solution was injected into the pronuclei of half-day B6CBF2 embryos, and 95 surviving microinjected embryos were transferred into the oviducts of pseudopregnant females. Twelve mice which developed from the microinjected embryos were born.

EXAMPLE 35

This example describes class-switching, somatic mutation, and B cell development in immunized transgenic mice homozygous for an inactivated endogenous immunoglobulin locus and containing the HC1 or HC2 heavy chain transgene(s).

To demonstrate that a human sequence germline configuration minilocus can functionally replace the authentic locus, we bred a mouse strain lacking endogenous IgH with strains containing human germline-configuration IgH transgenes. The two transgene miniloci, HC1 and HC2, include one and four functional variable (V) segments respectively 10 and 16 diversity (D) segments respectively, all six joining (JH) segments, and both the μ and $\gamma 1$ constant region segments. The miniloci include human cis-acting regulatory sequences--such as the JH- μ intronic enhancer and the μ and $\gamma 1$ switch sequences--that are closely linked to the coding segments. They also include an additional enhancer element derived from the 3' end of the rat IgH locus. We crossed HC1 and HC2 transgenic mice with stem-cell derived mutant mice that lack JH segments (JHD mice) as described (supra) and cannot therefore undergo functional heavy chain rearrangements. The resulting transgenic-JHD mice contain B cells that are dependent on the introduced heavy chain sequences.

Immunizations and hybridomas.

We immunized mice by intraperitoneal injections of 50-100 μ g of antigen. Antigens included human carcinoembryonic antigen (CEA; Crystal Chem, Chicago, IL), hen eggwhite lysozyme (HEL; Pierce, Rockford, IL), and keyhole limpet hemocyanin (KLH; Pierce, Rockford, IL). For primary injections we mixed the antigen with complete Freund's adjuvant, for subsequent injections we used incomplete Freund's adjuvant (Gibco BRL, Gaithersburg, MD). We fused spleen cells with the non-secreting mouse myeloma P3X63-Ag8.653 (ATCC, CRL1580). We assayed serum samples and hybridoma supernatants for the presence of specific and non-specific antibody comprising human heavy chain sequences by ELISA. For detection of non-specific antibodies we coated microtiter wells with human heavy chain isotype specific antibody (mouse MAb α human IgG1, clone HP6069, Calbiochem, La Jolla, CA; mouse MAb α human IgM, clone CH6, The Binding Site, Birmingham, UK) and developed with peroxidase conjugated antisera (horseradish peroxidase conjugated affinity purified

206

fab fragment from polyclonal goat α human IgG(fc), cat # 109-036-098; affinity purified horseradish peroxidase conjugated polyclonal rabbit α human IgM(fc), cat # 309-035-095. Jackson Immuno Research, West Grove, PA). For detection of antigen-specific antibodies we coated microtiter wells with antigen and developed with peroxidase-conjugated human heavy chain isotype specific antisera. We detected bound peroxidase by incubation with hydrogen peroxide and 2,2'-Azino-bis-(3-Ethylbenzthiazoline-6-Sulfonic Acid, Sigma Chem. Co., St. Louis, MO). The reaction product is measured by absorption at 415 nm, and corrected for absorption at 490 nm.

Flow cytometry.

We prepared single cell suspensions from spleen, bone marrow, and peritoneal cavity, and lysed red cells with NH_4Cl , as described by Mishell and Shiigi. The lymphocytes are stained with the following reagents: Phycoerythrin conjugated anti-mouse Igk (clone X36; Becton Dickinson, San Jose, CA), FITC conjugated anti-mouse IgD (clone SBA 1, Southern Biotech, AL), FITC conjugated anti-mouse CD5 (clone 53-7.3; Becton Dickinson, San Jose, CA), FITC conjugated anti-mouse Ig λ (clone R26-46; Pharmingen, San Diego, CA), and Cy-Chrome conjugated anti-mouse B220 (clone RA3-6B2; Pharmingen, San Diego, CA). We analyzed the stained cells using a FACScan flow cytometer and LYSIS II software (Becton Dickinson, San Jose, CA). Most macrophages, neutrophils, and residual red cells are excluded by gating on forward and side scatter.

Rescue of B cell compartment

In the peritoneal cavity of HCl transgenic-JHD animals we find normal levels of CD5^+ B cells and approximately one-quarter the normal level of conventional CD5^- B cells. The transgenic peritoneal CD5^+ B cells are similar to the so-called B-1 cells described in normal animals: they are larger than conventional B and T lymphocytes, they express lower levels of B220 than the conventional B cells found in the spleen, and they include a higher proportion of λ light chain expressing

cells. Over 90% of the splenic B cells express κ , while up to 50% of the peritoneal B cells express λ . Thus, while the level of conventional B cells is uniformly reduced in all tissues, the level of B-1, which are reported to have a much greater capacity for self-renewal, appears to be normal in the HCl transgenic-JHD animals.

Class switching.

In transgenic-JHD mice, repeated exposure to antigen results in the production of human $\gamma 1$ antibodies as well as μ antibodies. We injected human CEA into transgenic-JHD mice at weekly intervals and monitored the serum levels of antigen-specific IgM and IgG1 over a period of four weeks (Fig. 63). At one week there is a detectable IgM response but no IgG1 response. However, the IgG1 response is greater than the IgM response after two weeks, and it continues to increase while the IgM response remains relatively constant. This pattern--an initial IgM reaction followed by an IgG reaction--is typical of a secondary immune response; and it suggests that cis-acting sequences included in the transgene may be responding to cytokines that direct class switching. We have considered three possible mechanisms for expression of non- μ isotypes, each of which have been discussed in the literature. These mechanisms are: alternative splicing, which does not involve deletion of the μ gene; " δ -type" switching, which involved deletion of the μ gene via homologous recombination between flanking repeat sequences; and non-homologous recombination between switch regions. The results of our experiments, described below, are indicative of a switch region recombination model.

Two types of non-deletional alternative splicing mechanisms can be invoked to explain an isotype shift. First, it is possible that a single transcript covering both μ and $\gamma 1$ is expressed from the transgene; this transcript could be alternatively spliced in response to cytokines induced by exposure to antigen. Alternative, a cytokine induced sterile transcript initiating upstream of $\gamma 1$ could be trans-spliced to the μ transcript. If either of these mechanisms were

responsible for the expression of human $\gamma 1$ sequences, then we would expect to be able to isolate hybridomas that express both μ and $\gamma 1$. However, although we have screened several hundred hybridomas expressing either human μ or human $\gamma 1$, we have not found any such double producer (μ^+ , $\gamma 1^+$) hybridomas. This indicates that expression of $\gamma 1$ is accompanied by deletion of the μ gene.

Deletion of the μ gene can be mediated by non-homologous recombination between the μ and $\gamma 1$ switch regions, or by homologous recombination between the two flanking 400 bp direct repeats ($\sigma\mu$ and $\Sigma\mu$) that are included in the HC1 and HC2 transgenes. Deletional recombination between $\sigma\mu$ and $\Sigma\mu$ has been reported to be responsible for the IgD^+ , IgM^- phenotype of some human B cells. While the first mechanism, non-homologous switch recombination, should generate switch products of varying lengths, the second mechanism, $\sigma\mu/\Sigma\mu$ recombination, should always generate the same product. We performed a Southern blot analysis of genomic DNA isolated from three hybridomas (Fig. 64A), one expressing μ and two expressing $\gamma 1$. We find genomic rearrangements upstream of the transgene $\gamma 1$ only in the two the $\gamma 1$ switch regions (Fig. 64B). Furthermore, neither of the observed structures is compatible with homologous recombination between $\sigma\mu$ and $\Sigma\mu$. Our results are therefore consistent with a model for $\gamma 1$ isotype expression mediated by deletional non-homologous recombination between the transgene encoded μ and $\gamma 1$ switch regions.

Trans-switching.

In addition to human $\gamma 1$, we find mouse γ in the serum of HC1 and HC2 transgenic-JHD mice. We have also obtained mouse γ expressing hybridomas from these animals. Because the non-transgenic homozygous JHD animals do not express detectable levels of mouse immunoglobulins, we attribute the expression of mouse γ in the HC1 and HC2 transgenic-JHD animals to the phenomenon of trans-switching. All of the transgenic hybridomas that we have analyzed express either mouse or human constant region sequences, but not both. It is therefore unlikely that a trans-splicing mechanism is

involved. We used PCR amplification to isolate cDNA clones of trans-switch products, and determined the nucleotide sequence of 10 of the resulting clones (Fig. 65). The 5' oligonucleotide in the PCR amplification is specific for the transgene encoded VH251, and the 3' oligonucleotide is specific for mouse $\gamma 1$, $\gamma 2b$, and $\gamma 3$ sequences. We find examples of trans-switch products incorporating all three of these mouse constant regions.

Somatic mutation.

Approximately 1% of the nucleotides within the variable regions of the trans-switch products shown in Fig. 7 are not germline encoded. This is presumably due to somatic mutation. Because the mutated sequence has been translocated to the endogenous locus, the cis-acting sequences directing these mutations could be located anywhere 3' of the mouse γ switch. However, as we discuss below, we also observe somatic mutation in VDJ segments that have not undergone such translocations; and this result indicates that sequences required by heavy chain somatic mutation are included in the transgene.

To determine if the HC1 and HC2 constructs include sufficient cis-acting sequences for somatic mutation to occur in the transgenic-JHD mice, we isolated and partially sequenced cDNA clones derived from two independent HC1 transgenic lines and one HC2 line. We find that some of the $\gamma 1$ transcripts from transgenic-JHD mice contain V regions with extensive somatic mutations. The frequency of these mutated transcripts appears to increase with repeated immunizations. Figs. 66A and 66B show two sets of cDNA sequences: one set is derived from an HC1 (line 26) transgenic-JHD mouse that we immunized with a single injection of antigen 5 days before we isolated RNA; the second set is derived from an HC1 (line 26) transgenic-JHD mouse that we hyperimmunized by injecting antigen on three different days beginning 5 months before we isolated RNA; the second set is derived from an HC1 (line 26) transgenic-JHD mouse that we hyperimmunized by injecting antigen on three different days beginning 5 months before we

isolated RNA. Only 2 of the 13 V regions from the 5 day post-exposure mouse contain any non-germline encoded nucleotides. Each of these V's contains only a single nucleotide change, giving an overall somatic mutation frequency of less than 0.1% for this sample. In contrast, none of the 13 V sequences from the hyperimmunized animal are completely germline, and the overall somatic mutation frequency is 1.6%.

Comparison of μ and $\gamma 1$ transcripts isolated from a single tissue sample shows that the frequency of somatic mutations is higher in transgene copies that have undergone a class switch. We isolated and partially sequenced 47 independent μ and $\gamma 1$ cDNA clones from a hyperimmunized CH1 line 57 transgenic-JHD mouse (Fig. 67A and 67B). Most of the μ cDNA clones are unmodified relative to the germline sequence, while over half of the $\gamma 1$ clones contain multiple non-germline encoded nucleotides. The $\gamma 1$ expressing cells are distinct from the μ expressing cells and, while the two processes are not necessarily linked, class switching and somatic mutation are taking place in the same sub-population of B cells.

Although we do not find extensive somatic mutation of the VH251 gene in non-hyperimmunized CH1 transgenic mice, we have found considerable somatic mutation in VH56p1 and VH51p1 genes in a naive HC2 transgenic mouse. We isolated spleen and lymph node RNA from an unimmunized 9 week old female HC2 transgenic animal. We individually amplified $\gamma 1$ transcripts that incorporate each of the four V regions in the HC2 transgene using V and $\gamma 1$ specific primers. The relative yields of each of the specific PCR products were VH56p1>>VH51p1>VH4.21>VH251. Although this technique is not strictly quantitative, it may indicate a bias in V segment usage in the HC2 mouse. Fig. 68 shows 23 randomly picked $\gamma 1$ cDNA sequences derived from PCR amplifications using an equimolar mix of all four V specific primers. Again we observe a bias toward VH56p1 (19/23 clones). In addition, the VH56p1 sequences show considerable somatic mutation, with an overall frequency of 2.1% within the V gene segment. Inspection of the CDR3 sequences reveals that although 17 of

the 19 individual VH56p1 clones are unique, they are derived from only 7 different VDJ recombination events. It thus appears that the VH56p1 expressing B cells are selected, perhaps by an endogenous pathogen or self antigen, in the naive animal. It may be relevant that this same gene is over-represented in the human fetal repertoire.

Summary

Upstream cis-acting sequences define the functionality of the individual switch regions, and are necessary for class switching. Our observation--that class switching within the HC1 transgene is largely confined to cells involved in secondary response, and does not occur randomly across the entire B cell population--suggests that the minimal sequences contained within the transgene are sufficient. Because the γ sequences included in this construct begin only 116 nucleotides upstream of the start site of the γ 1 sterile transcript, the switch regulatory region is compact.

Our results demonstrate that these important cis-acting regulatory elements are either closely linked to individual γ genes, or associated with the 3' heavy chain enhancer included in the HC1 and HC2 transgenes. Because the HC1 and HC2 inserts undergo transgene-autonomous class switching--which can serve as a marker for sequences that are likely to have been somatically mutated--we were able to easily find hypermutated transcripts that did not originate from translocations to the endogenous locus. We found somatically mutated γ transcripts in three independent transgenic lines (two HC1 lines and one HC2 line). It is therefore unlikely that sequences flanking the integration sites of the transgene affect this process; instead, the transgene sequences are sufficient to direct somatic mutation.

EXAMPLE 36

This example describes the generation of hybridomas from mice homozygous for an inactivated endogenous immunoglobulin locus and containing transgene sequences

encoding a human sequence heavy chain and human sequence light chain. The hybridomas described secrete monoclonal antibodies comprising a human sequence heavy chain and a human sequence light chain and bind to a predetermined antigen expressed on T lymphocytes. The example also demonstrates the capacity of the mice to make a human sequence antibody in response to a human-derived immunogen, human CD4, and the suitability of such mice as a source for making hybridomas secreting human sequence monoclonal antibodies reactive with human antigens.

A. Generation of Human Ig Monoclonal Antibodies Derived from HC1 Transgenic Mice Immunized with a Human CD4 Antigen

A transgenic mouse homozygous for a functionally disrupted J_H locus and harboring a transgene capable of rearranging to encode a human sequence heavy chain and a transgene capable of rearranging to encode a human sequence light chain was immunized. The genotype of the mouse was HC1-26⁺ KC1e-1536⁺ $J_H D^{+/+}$ $J_K D^{-}$, indicating homozygosity for murine heavy chain inactivation and the presence of germline copies of the HC1 human sequence heavy chain transgene and the KC1e human sequence light chain transgene.

The mouse was immunized with a variant of the EL4 cell line (ATCC) expressing a mouse-human hybrid CD4 molecule encoded by a stably transfected polynucleotide. The expressed CD4 molecule comprises a substantially human-like CD4 sequence. Approximately 5×10^6 cells in 100 μ l of PBS accompanied by 100 μ l of Complete Freund's Adjuvant (CFA) were introduced into the mouse via intraperitoneal injection on Day 0. The inoculation was repeated on Days 7, 14, 21, 28, 60, and 77, with test bleeds on Days 18, 35, and 67. The spleen was removed on Day 81 and approximately 7.2×10^7 spleen cells were fused to approximately 1.2×10^7 fusion partner cells (P3x63Ag8.653 cell line; ATCC) by standard methods (PEG fusion) and cultured in RPMI 1640 15 % FCS, 4 mM glutamine, 1 mM sodium pyruvate plus HAT and PSN medium. Multiple fusions were performed.

Hybridomas were grown up and supernatants were tested with ELISA for binding to a commercial source of

purified recombinant soluble human sequence CD4 expressed in CHO cells (American Bio-Technologies, Inc. (ABT), Cambridge, MA) and/or CD4 obtained from NEN-DuPont. The ABT sample contained a purified 55 kD human CD4 molecule comprised the V₁ through V₃ domains of human CD4. The recombinant human sequence CD4 (produced in CHO-K1 cells) was adsorbed to the assay plate and used to capture antibody from hybridoma supernatants, the captured antibodies were then evaluated for binding to a panel of antibodies which bind either human μ , human κ , human γ , murine μ , or murine κ .

One hybridoma was subcloned from its culture plate well, designated 1F2. The 1F2 antibody bound to the ABT CD4 preparation, was positive for human μ and human κ , and was negative for human γ , mouse γ , and mouse κ .

B. Generation of Human Ig Monoclonal Antibodies Derived from HC2 Transgenic Mice Immunized with Human CD4 and Human IgE.

The heavy chain transgene, HC2, is shown in Fig. 56 and has been described supra (see, Example 34).

The human light chain transgene, KCo4, depicted in Fig. 56 is generated by the cointegration of two individually cloned DNA fragments at a single site in the mouse genome. The fragments comprise 4 functional V κ segments, 5J segments, the C κ exon, and both the intronic and downstream enhancer elements (see Example 21) (Meyer and Neuberger (1989), EMBO J. 8:1959-1964; Judde and Max (1992), Mol. Cell Biol. 12:5206-5216). Because the two fragments share a common 3 kb sequence (see Fig. 56), they can potentially integrate into genomic DNA as a contiguous 43 kb transgene, following homologous recombination between the overlapping sequences. It has been demonstrated that such recombination events frequently occur upon microinjection of overlapping DNA fragments (Pieper et al. (1992), Nucleic Acids Res. 20:1259-1264). Co-injected DNA's also tend to co-integrate in the zygote, and the sequences contained within the individually cloned fragments would subsequently be jointed by DNA rearrangement during B cell development. Table 11 ~~12~~ shows that transgene inserts from at least 2 of the transgenic lines are functional.

Examples of VJ junctions incorporating each of the 4 transgene encoded V segments, and each of the 5J segments, are represented in this set of 36 clones.

Table 11 12

TP160

line	V _k 65.5	V _k 65.8	V _k 65.15	V _k 65.3	J _k 1	J _k 2	J _k 3	J _k 4	J _k 5
#4436	0	11	4	3	14	1	0	2	1
#4437	1	3	7	7	5	2	1	7	3

Human light chain V and J segment usage in KCo4 transgenic mice. The table shows the number of PCR clones, amplified from cDNA derived from two transgenic lines, which contain the indicated human kappa sequences. cDNA was synthesized using spleen RNA isolated from individual KCo4 transgenic mice (mouse #8490, 3 mo., male, KCo4 line 4437; mouse #8867, 2.5 mo., female, KCo4 line 4436). The cDNA was amplified by PCR using a C_k specific oligonucleotide, 5' TAG AAG GAA TTC AGC AGG CAC ACA ACA GAG GCA GTT CCA 3' (SEQ ID NO:173), and a 1:3 mixture of the following 2 V_k specific oligonucleotides: 5' AGC TTC TCG AGC TCC TGC TGC TCT GTT TCC CAG GTG CC 3' (SEQ ID NO:174) and 5' CAG CTT CTC GAG CTC CTG CTA CTC TGG CTC (C,A)CA GAT ACC 3' (SEQ ID NO:175). The PCR product was digested with XhoI and EcoRI, and cloned into a plasmid vector. Partial nucleotide sequences were determined by the dideoxy chain termination method for 18 randomly picked clones from each animal. The sequences of each clone were compared to the germline sequence of the unarranged transgene.

Twenty-three light chain minilocus positive and 18 heavy chain positive mice developed from the injected embryos. These mice, and their progeny, were bred with mice containing targeted mutations in the endogenous mouse heavy (strain JHD) and κ light chain loci (strain JCKD) to obtain mice containing human heavy and κ light chain in the absence of functional mouse heavy and κ light chain loci. In these mice, the only mouse light chain contribution, if any, is from the mouse λ locus.

Table 12 shows that somatic mutation occurs in the variable regions of the transgene-encoded human heavy chain transcripts of the transgenic mice. Twenty-three cDNA clones from a HC2 transgenic mouse were partially sequenced to determine the frequency of non-germline encoded nucleotides within the variable region. The data include only the sequence of V segment codons 17-94 from each clone, and does not include N regions. RNA was isolated from the spleen and lymph node of mouse 5250 (HC2 line 2550 hemizygous, JHD homozygous). Single-stranded cDNA was synthesized and γ transcripts amplified by PCR as described [references]. The amplified cDNA was cloned into plasmid vectors, and 23 randomly picked clones were partially sequenced by the dideoxy chain-termination method. The frequency of PCR-introduced nucleotide changes is estimated from constant region sequence as <0.2%.

TABLE 12: The Variable Regions of Human γ Transcripts in HC2 Transgenic Mice Contain Non-Germline-Encoded Nucleotides

VH Segment	Number of clones	Number of non-germline encoded nucleotides	Frequency of non-germline-encoded nucleotides (%)
VH251	0		--
VH56P1	10	100	2.1
VH51P1	1	5	2.0
VH4.21	3	0	0.0

Flow cytometry

We analyzed the stained cells using a FACScan flow cytometer and LYSIS II software (Becton Dickinson, San Jose, CA). Spleen cells were stained with the following reagents: propidium iodide (Molecular Probes, Eugene, OR), phycoerythrin conjugated α -human Igk (clone HP6062; Caltag, S. San Francisco, CA), phycoerythrin conjugated α -mouse Igk (clone X36; Becton Dickinson, San Jose, CA), FITC conjugated α -mouse Ig λ (clone R26-46; Pharmingen, San Diego, CA), FITC conjugated α -mouse Ig μ (clone R6-60.2; Pharmingen, San Diego, CA), FITC conjugated α -human Ig μ (clone G20-127; Pharmingen, San Diego, CA), and Cy-Chrome conjugated α -mouse B220 (clone RA3-6B2; Pharmingen, San Diego, CA).

Expression of human Ig transgenes

Figure 69 shows a flow cytometric analysis of spleen cells from KCo4 and HC2 mice that are homozygous for both the JHD and JCKD mutations. The human sequence HC2 transgene rescued B cell development in the JHD mutant background, restoring the relative number of B220⁺ cells in the spleen to approximately half that of a wild type animal. These B cells expressed cell surface immunoglobulin receptors that used transgene encoded heavy chain. The human KCo4 transgene was also functional, and competed successfully with the intact endogenous λ light chain locus. Nearly 95% of the splenic B cells in JHD/JCKD homozygous mutant mice that contain both heavy and light chain human transgenes (double transgenic) expressed completely human cell surface IgMk.

Serum Ig levels were determined by ELISA done as follows: human μ : microtiter wells coated with mouse Mab α human IgM (clone CH6, The Binding Site, Birmingham, UK) and developed with peroxidase conjugated rabbit α human IgM(fc) (cat # 309-035-095, Jackson Immuno Research, West Grove, PA). Human γ : microtiter wells coated with mouse Mab α human IgG1 (clone HP6069, Calbiochem, La Jolla, CA) and developed with peroxidase conjugated goat α human IgG(fc) (cat # 109-036-098, Jackson Immuno Research, West Grove, PA). Human κ : microtiter wells coated with mouse Mab α human Igk (cat # 0173, AMAC, Inc. Igk (cat #A7164, Sigma Chem. Co., St. Louis, MO). Mouse γ :

microtiter wells coated with goat α mouse IgG (cat #115-006-071, Jackson Immuno Research, West Grove, PA). Mouse λ : microtiter wells coated with rat MAb α mouse Ig λ (cat # 02171D, Pharmingen, San Diego, CA) and developed with peroxidase conjugated rabbit α mouse IgM(fc) (cat # 309-035-095, Jackson Immuno Research, West Grove, PA). Bound peroxidase is detected by incubation with hydrogen peroxide and 2,2'-Azino-bis-(3-Ethylbenzthiazoline-6-Sulfonic Acid, Sigma Chem. Co., St. Louis, MO). The reaction product is measured by absorption at 415 nm.

The double transgenic mice also express fully human antibodies in the serum. Figure 70 shows measured serum levels of immunoglobulin proteins for 18 individual double transgenic mice, homozygous for endogenous heavy and kappa light chain inactivations, derived from several different transgenic founder animals. We found detectable levels of human μ , γ 1, and κ . We have shown supra that the expressed human γ 1 results from authentic class switching by genomic recombination between the transgene μ and γ 1 switch regions. Furthermore, we have found that intra-transgene class switching was accompanied by somatic mutation of the heavy chain variable regions. In addition to human immunoglobulins, we also found mouse γ and λ in the serum. The present of mouse λ protein is expected because the endogenous locus is completely intact. We have shown elsewhere that the mouse γ expression is a consequence of trans-switch recombination of transgene VDJ segments into the endogenous heavy chain locus. This trans-switching phenomenon, which was originally demonstrated for wild-type heavy chain alleles and rearranged VDJ transgenes (Durdik et al. (1989), Proc. Natl. Acad. Sci. USA 86:2346-2350; Gerstein et al. (1990), Cell 63:537-548), occurs in the mutant JHD background because the downstream heavy chain constant regions and their respective switch elements are still intact.

The serum concentration of human IgM κ in the double transgenic mice was approximately 0.1 mg/ml, with very little deviation between animals or between lines. However, human γ 1, mouse γ , and mouse λ levels range from 0.1 to 10 micrograms/ml. The observed variation in γ levels between individual animals

may be a consequence of the fact that γ is an inducible constant region. Expression presumably depends on factors such as the health of the animal, exposure to antigens, and possibly MHC type. The mouse λ serum levels are the only parameter that appears to correlate with individual transgenic lines. KCo4 line 4436 mice which have the fewest number of copies of the transgene per integration (approximately 1-2 copies) have the highest endogenous λ levels, while KCo4 line 4437 mice (~10 copies per integration) have the lowest λ levels. This is consistent with a model in which endogenous λ rearranges subsequent to the κ transgene, and in which the serum λ level is not selected for, but is instead a reflection of the relative size of the precursor B cell pool. Transgene loci containing multiple light chain inserts may have the opportunity to undergo more than one V to J recombination event, with an increased probability that one of them will be functional. Thus high copy lines will have a smaller pool of potential λ cells.

Immunizations with human CD4 and IgE

To test the ability of the transgenic B cells to participate in an immune response, we immunized double transgenic mice with human protein antigens, and measured serum levels of antigen specific immunoglobulins by ELISA. Mice were immunized with 50 μ g recombinant sCD4 (cat. # 013101, American Bio-Technologies Inc., Cambridge, MA) covalently linked to polystyrene beads (cat # 08226, Polysciences Inc., Warrington, PA) in complete Freund's adjuvant by intraperitoneal injection. Each of the mice are homozygous for disruptions of the endogenous μ and κ loci, and hemizygous for the human heavy chain transgene HC2 line 2500 and human κ light chain transgene KCo4 line 4437.

Methods

Serum samples were diluted into microtiter wells coated with recombinant sCD4. Human antibodies were detected with peroxidase conjugated rabbit α human IgM(fc) (Jackson Immuno Research, West Grove, PA) or peroxidase conjugated goat anti-human Igk (Sigma, St. Louis, MO).

Figure 71A shows the primary response of transgenic mice immunized with recombinant human soluble CD4. All four of the immunized animals show an antigen-specific human IgM response at one week. The CD4-specific serum antibodies comprise both human μ heavy chain and human κ light chain.

To evaluate the ability of the HC2 transgene to participate in a secondary response, we hyperimmunized the transgenic mice by repeated injection with antigen, and monitored the heavy chain isotype of the induced antibodies. Mice homozygous for the human heavy chain transgene HC2 and human κ light chain transgene KCo4 were immunized with 25 μ g of human IgE_K (The Binding Site, Birmingham, UK) in complete Freund's adjuvant on day = 0. Thereafter, animals were injected with IgE_K in incomplete Freund's adjuvant at approximately weekly intervals. Serum samples were diluted 1:10, and antigen-specific ELISAs were performed on human IgE, λ coated plates.

Figure 71B shows a typical time course of the immune response from these animals: we injected double transgenic mice with human IgE in complete Freund's adjuvant, followed by weekly boosts of IgE in incomplete Freund's adjuvant. The initial human antibody response was IgM_K, followed by the appearance of antigen specific human IgG_K. The induced serum antibodies in these mice showed no cross-reactivity to human IgM or BSA. The development, over time, of a human IgG

We have also tested the ability of the heavy chain transgene to undergo class switching *in vitro*: splenic B cells purified from animals hemizygous for the same heavy chain construct (HC2, line 2550) switch from human IgM to human IgG1 in the presence of LPS and recombinant mouse IL-4. However, *in vitro* switching did not take place in the presence of LPS and recombinant mouse IL-2, or LPS alone.

We find human IgM-expressing cells in the spleen, lymph nodes, peritoneum, and bone marrow of the double-transgenic/double-knockout (0011) mice. Although the peritoneal cavity contains the normal number of B cells, the absolute number of transgenic B cells in the bone marrow and spleen is approximately 10-50% of normal. The reduction may

result from a retardation in transgene-dependent B cell development. The double-transgenic/double-knockout (0011) mice also express fully human antibodies in the serum, with significant levels of human μ , $\gamma 1$, and κ in these mice. The expressed human $\gamma 1$ results from authentic class switching by genomic recombination between the transgene μ and $\gamma 1$ switch regions. Furthermore, the intratransgene class switching is accompanied by somatic mutation of the heavy chain variable regions encoded by the transgene. In addition to human immunoglobulins, we find mouse μ and mouse λ in these mice. The mouse μ expression appears to be a result of trans-switching recombination, wherein transgene VDJ gene is recombined into the endogenous mouse heavy chain locus. Trans-switching, which was originally observed in the literature for wild-type heavy chain alleles and rearranged VDJ transgenes, occurs in our $J_H^{-/-}$ background because the mouse downstream heavy chain constant regions and their respective switch elements are still intact.

To demonstrate the ability of the transgenic B cells to participate in an immune response, we immunized the 0011 mice with human protein antigens, and monitored serum levels of antigen-specific immunoglobulins. The initial human antibody response is IgM, followed by the expression of antigen-specific human IgG (Fig. 71B and Fig. 73). The lag before appearance of human IgG antibodies is consistent with an association between class-switching and a secondary response to antigen.

In a transgenic mouse immunized with human CD4, human IgG reactivity to the CD4 antigen was detectable at serum concentrations ranging from 2×10^{-2} to 1.6×10^{-4} .

Identification of Anti-Human CD4 Hybridomas

A transgenic mouse homozygous for the human heavy chain transgene HC2 and human κ light chain transgene KCo4 were immunized with 20 μ g of recombinant human CD4 in complete Freund's adjuvant on day 0. Thereafter, animals were injected with CD4 in incomplete Freund's adjuvant at approximately weekly intervals. Fig. 73 shows human antibody response to human CD4 in serum of the transgenic mouse. Serum samples were

diluted 1:50, and antigen-specific ELISAs were performed on human CD4 coated plates. Each line represents individual sample determinations. Solid circles represent IgM, open squares represent IgG.

We also isolated hybridoma cell lines from one of the mice that responded to human CD4 immunization. Five of the cloned hybridomas secrete human IgGk (human $\gamma 1$ /human κ) antibodies that bind to recombinant human CD4 and do not crossreact (as measured by ELISA) with a panel of other glycoprotein antigens. The association and dissociation rates of the immunizing human CD4 antigen for the monoclonal antibodies secreted by two of the IgGk hybridomas, 4E4.2 and 2C5.1, were determined. The experimentally-derived binding constants (K_a) were approximately $9 \times 10^7 \text{ M}^{-1}$ and $8 \times 10^7 \text{ M}^{-1}$ for antibodies 4E4.2 and 2C5.1, respectively. These K_a values fall within the range of murine IgG anti-human CD4 antibodies that have been used in clinical trials by others (Chen et al. (1993) Int. Immunol. 6: 647).

A mouse of line #7494 (0012;HC1-26+;JHD++;JKD++;KC2-1610++) was immunized on days 0, 13, 20, 28, 33, and 47 with human CD4, and produced anti-human CD4 antibodies comprised of human κ and human μ or γ .

By day 28, human μ and human κ were found present in the serum. By day 47, the serum response against human CD4 comprised both human μ and human γ , as well as human κ . On day 50, splenocytes were fused with P3X63-Ag8.653 mouse myeloma cells and cultured. Forty-four out of 700 wells (6.3%) contained human γ and/or κ anti-human CD4 monoclonal antibodies. Three of these wells were confirmed to contain human γ anti-CD4 monoclonal antibodies, but lacked human κ chains (presumably expressing mouse λ). Nine of the primary wells contained fully human IgMk anti-CD4 monoclonal antibodies, and were selected for further characterization. One such hybridoma expressing fully human IgMk anti-CD4 monoclonal antibodies was designated 2C11-8.

Primary hybridomas were cloned by limiting dilution and assessed for secretion of human μ and κ monoclonal antibodies reactive against CD4. Five of the nine hybridomas

remained positive in the CD4 ELISA. The specificity of these human IgMk monoclonal antibodies for human CD4 was demonstrated by their lack of reactivity with other antigens including ovalbumin, bovine serum albumin, human serum albumin, keyhole limpet hemacyanin, and carcinoembryonic antigen. To determine whether these monoclonal antibodies could recognize CD4 on the surface of cells (i.e., native CD4), supernatants from these five clones were also tested for reactivity with a CD4+ T cell line, Sup T1. Four of the five human IgMk monoclonal antibodies reacted with these CD4+ cells. To further confirm the specificity of these IgMk monoclonal antibodies, freshly isolated human peripheral blood lymphocytes (PBL) were stained with these antibodies. Supernatants from clones derived from four of the five primary hybrids bound only to CD4+ lymphocytes and not to CD8+ lymphocytes (Figure 72).

Fig. 72 shows reactivity of IgMk anti-CD4 monoclonal antibody with human PBL. Human PBL were incubated with supernatant from each clone or with an isotype matched negative control monoclonal antibody, followed by either a mouse anti-human CD4 monoclonal antibody conjugated to PE (top row) or a mouse anti-human CD8 Ab conjugated to FITC (bottom row). Any bound human IgMk was detected with a mouse anti-human μ conjugated to FITC or to PE, respectively. Representative results for one of the clones, 2C11-8 (right side) and for the control IgMk (left side) are shown. As expected, the negative control IgMk did not react with T cells and the goat anti-human μ reacted with approximately 10% of PBL, which were presumably human B cells.

Good growth and high levels of IgMk anti-CD4 monoclonal antibody production are important factors in choosing a clonal hybridoma cell line for development. Data from one of the hybridomas, 2C11-8, shows that up to 5 pg/cell/d can be produced (Figure 74). Similar results were seen with a second clone. As is commonly observed, production increases dramatically as cells enter stationary phase growth.

Fig. 74 shows cell growth and human IgMk anti-CD4 monoclonal antibody secretion in small scale cultures. Replicate cultures were seeded at 2×10^5 cells/ml in a total

volume of 2 ml. Every twenty-four hours thereafter for four days, cultures were harvested. Cell growth was determined by counting viable cells and IgMk production was quantitated by an ELISA for total human μ (top panel). The production per cell per day was calculated by dividing the amount of IgMk by the cell number (bottom panel).

Fig. 75 shows epitope mapping of a human IgMk anti-CD4 monoclonal antibody. Competition binding flow cytometric experiments were used to localize the epitope recognized by the IgMk anti-CD4 monoclonal antibody, 2C11-8. For these studies, the mouse anti-CD4 monoclonal antibodies, Leu3a and RPA-T4, which bind to unique, nonoverlapping epitopes on CD4 were used. PE fluorescence of CD4⁺ cells preincubated with decreasing concentrations of either RPA-TA or Leu-3a followed by staining with 2C11-8 detected with PE-conjugated goat anti-human IgM. There was concentration-dependent competition for the binding of the human IgMk anti-CD4 monoclonal antibody 2C11-8 by Leu3a but not by RPA-T4 (Figure 75). Thus, the epitope recognized by 2C11-8 was similar to or identical with that recognized by monoclonal antibody Leu3a, but distinct from that recognized by RPA-T4.

In summary, we have produced several hybridoma clones that secrete human IgMk monoclonal antibodies that specifically react with native human CD4 and can be used to discriminate human PBLs into CD4⁺ and CD4⁻ subpopulations. At least one of these antibodies binds at or near the epitope defined by monoclonal antibody Leu3a. Monoclonal antibodies directed to this epitope have been shown to inhibit a mixed leukocyte response (Engleman et al., J. Exp. Med. (1981) 153:193). A chimeric version of monoclonal antibody Leu3a has shown some clinical efficacy in patients with mycosis fungoides (Knox et al. (1991) Blood 77:20).

We have isolated cDNA clones from 3 different hybridoma cell lines (2C11.8, 2C5.1, and 4E4.2), and have determined the partial nucleotide sequence of some of the expressed immunoglobulin genes in each of these cell lines.

For sequence analysis, total RNA was isolated from approximately 5×10^6 hybridoma cells. ssDNA was synthesized

by priming reverse transcription with oligo dT. A portion of this ssDNA was used in duplicate PCR reactions primed by a pool of oligos with specificities for either (i) heavy chain variable framework regions contained within the HC1 or HC2 transgenes and a single downstream oligo specific for constant human gamma sequence, or (ii) light chain variable framework regions contained within the KC2 or KCo4 transgene and a single downstream oligo specific for constant human kappa sequence. Products from these PCR reactions were digested with appropriate restriction enzymes, gel purified, and independently cloned into pNNO3 vector. DNA was isolated and manual dideoxy and/or automated fluorescent sequencing reactions performed on dsDNA.

The characteristics of the three hybridomas, 2C11.8, 2C5.1, and 4E4.2, are given below in Table 11.

Table 13 Human variable region usage in hybridomas

Subclone	Specificity	Isotype	Vh	Dh	Jh	Vk	Jk
2C11.8	nCD4	IgM κ	251	nd.*	nd.	nd.	nd.
2C5.1	rCD4	IgG κ	251	HQ52	JHS	65.15	JK4
4E4.2	rCD4	IgG κ	251	HQ52	JHS	65.15	JK4

* n.d., not determined

Nucleotide sequence analysis of expressed heavy and light chain sequences from the two IgG κ hybridomas 2C5.1 and 4E4.2 reveal that they are sibling clones derived from the same progenitor B cell. The heavy and light chain V(D)J junctions from the two clones are identical, although the precise nucleotide sequences differ by presumptive somatic mutations. The heavy chain VDJ junction sequence is:

VH251					N					DHQ52					JH5				
TAT	TAC	TGT	GCG	AG	(g	gct	cc)	A	ACT	GGG	GA	C	TGG	TTC	GAC				
Y	Y	C	A	R	A	P	T	G	D	W	F	D							

(SEQ ID NOS:176 and 177, respectively)

The light chain VJ junction is:

1,2270

	Vk65.15						N	Jk4		
	TAT	AAT	AGT	TAC	CCT	CC	(t)	ACT	TTC	GGC
	Y	N	S	Y	P	P		T	F	G

(SEQ ID NOS:178 and 179, respectively)

The following non-germline encoded codons were identified (presumptive somatic mutations):

2CS.1 heavy chain AGC->AGG S28R (replacement)
light chain CCG->ACG P119T (replacement)

4E4.2 heavy chain AGC->AGG S28R (replacement)
CTG->CTA L80L (silent)
light chain GAG->GAC E41D (replacement)
AGG->AAG R61K (replacement)
CCG->ACG P119T (replacement)

1,2271

We conclude that these two gamma hybridomas are derived from B cells that have undergone a limited amount of somatic mutation. This data shows that the HC2 transgenic animals use the VH5-51 (aka VH251) V segment. We have previously shown that VH4-34, VH1-69, and VH3-30.3 are expressed by these mice. The combination of these results demonstrates that the HC2 transgenic mice express all four of the transgene encoded human VH genes.

We conclude that human immunoglobulin-expressing B cells undergo development and respond to antigen in the context of a mouse immune system. Antigen responsivity leads to immunoglobulin heavy chain isotype switching and variable region somatic mutation. We have also demonstrated that conventional hybridoma technology can be used to obtain monoclonal human sequence antibodies from these mice. Therefore, these transgenic mice represent a source of human antibodies against human target antigens.

This example describes the generation of transgenic mice homozygous for an inactivated endogenous heavy chain and κ chain locus and harboring a transgene capable of isotype switching to multiple downstream human C_H genes. The example also demonstrates a cloning strategy for assembling large transgenes (e.g., 160 kb) by co-microinjection of multiple DNA fragments comprising overlapping homologous sequence joints (see Fig. 76), permitting construction of a large transgene from more than two overlapping fragments by homologous recombination of a plurality of homology regions at distal ends of the set of fragments to be assembled in vivo, such as in a microinjected ES cell or its clonal progeny. The example also shows, among other things, that isolated lymphocytes from the transgenic animals can be induced to undergo isotype switching in vitro, such as with IL-4 and LPS.

A set of five different plasmid clones was constructed such that the plasmid inserts could be isolated, substantially free of vector sequences; and such that the inserts together form a single imbricate set of overlapping sequence spanning approximately 150 kb in length. This set includes human V, D, J, μ , $\gamma 3$, and $\gamma 1$ coding sequences, as well as a mouse heavy chain 3' enhancer sequence. The five clones are, in 5' to 3' order: pH3V4D, pCOR1xa, p11-14, pP1-570, and pHP-3a (Fig. 76). Several different cloning vectors were used to generate this set of clones. Some of the vectors were designed specifically for the purpose of building large transgenes. These vectors (pGP1a, pGP1b, pGP1c, pGP1d, pGP1f, pGP2a, and pGP2b) are pBR322-based plasmids that are maintained at a lower copy number per cell than the pUC vectors (Yanisch-Perron et al. (1985) Gene 33: 103-119). The vectors also include trpA transcription termination signals between the polylinker and the 3' end of the plasmid β -lactamase gene. The polylinkers are flanked by restriction sites for the rare-cutting enzyme NotI; thus allowing for the isolation of the insert away from vector sequences prior to embryo microinjection. Inside of the NotI sites, the polylinkers include unique XhoI and SalI sites at either end. The pGP1 vectors are described in Taylor et al. (1992) Nucleic Acids

Res. 23: 6287. To generate the pGP2 vectors, pGP1f was first digested with AlwNI and ligated with the synthetic oligonucleotides o-236 and o-237 (o-236, 5' - ggc gcg cct tgg cct aag agg cca - 3' (SEQ ID NO:180); o-237, 5' - cct ctt agg cca agg cgc gcc tgg - 3' (SEQ ID NO:181)) The resulting plasmid is called pGP2a. Plasmid pGP2a was then digested with KpnI and EcoRI, and ligated with the oligonucleotides o-288 and o-289 (o-288, 5' - aat tca gta tcg atg tgg tac - 3' (SEQ ID NO:182); o-289, 5' - cac atc gat act g - 3' (SEQ ID NO:183)) to create pGP2b (Figs. 77A and Fig. 77B).

The general scheme for transgene construction with the pGP plasmids is outlined in Fig. 78 (paths A and B). All of the component DNA fragments are first cloned individually in the same 5' to 3' orientation in pGP vectors. Insert NotI, XhoI and SalI sites are destroyed by oligonucleotide mutagenesis or if possible by partial digestion, polymerase fill-in, and blunt end ligation. This leaves only the polylinker derived XhoI and SalI sites at the 5' and 3' ends of each insert. Individual inserts can then be combined stepwise by the process of isolating XhoI/SalI fragments from one clone and inserting the isolated fragment into either the 5' XhoI or 3' SalI site of another clone (Fig. 78, path A). Transformants are then screened by filter hybridization with one or more insert fragments to obtain the assembled clone. Because XhoI/SalI joints cannot be cleaved with either enzyme, the resulting product maintains unique 5' XhoI and 3' SalI sites, and can be used in the step of the construction. A variation of this scheme is carried out using the vectors pGP2a and pGP2b (Fig. 78, path B). These plasmids includes an SfiI site between the ampicillin resistance gene and the plasmid origin of replication. By cutting with SfiI and XhoI or SalI, inserts can be isolated together with either the drug resistance sequence or the origin of replication. One SfiI/XhoI fragment is ligated to one SfiI/SalI fragment in each step of the synthesis. There are three advantages to this scheme: (i) background transformants are reduced because sequences from both fragments are required for plasmid replication in the presence of ampicillin; (ii) the ligation can only occur in a

single 5' to 3' orientation; and (iii) the SfiI ends are not self-compatible, and are not compatible with SalI or XhoI, thus reducing the level of non-productive ligation. The disadvantage of this scheme is that insert SfiI sites must be removed as well as NotI, XhoI, and SalI sites. These medium copy vectors are an improvement over the commonly used pUC derived cloning vectors. To compare the ability of these vectors to maintain large DNA inserts, a 43 kb XhoI fragment comprising the human JH/C μ region was ligated into the SalI site of pSP72 (Promega, Madison, WI), pUC19 (BRL, Grand Island, NY), and pGP1f. Transformant colonies were transferred to nitrocellulose and insert containing clones were selected by hybridization with radiolabeled probe. Positive clones were grown overnight in 3 ml media and DNA isolated: EcoRI digestion of the resulting DNA reveals that all the pSP72 and pUC19 derived clones deleted the insert (Fig. 79); however, 12 of the 18 pGP1f derived clones contained intact inserts. Both orientations are represented in these 12 clones.

The construction and isolation of the five clones (pH3V4D, pCOR1xa, p11-14, pP1-570, and pHP-3a) used to generate the HCo7 transgene is outlined below.

pH3V4D.

Germline configuration heavy chain variable gene segments were isolated from phage 1 genomic DNA libraries using synthetic oligonucleotide probes for VH1 and VH3 classes. The VH1 class probe was o-49:

5' - gtt aaa gag gat ttt att cac ccc tgt gtc ctc tcc aca ggt
gtc - 3' (SEQ ID NO:78)

The VH3 class probe was o-184:

5' - gtt tgc agg tgt cca gtg t(c,g)a ggt gca gct g(g,t)t gga
gtc (t,c)(g,c)g - 3' (SEQ ID NO:184).

Positively hybridizing clones were isolated, partially restriction mapped, subcloned and partially sequenced. From the nucleotide sequence it was determined that one of the VH1 clones isolated with the o-49 probe encoded a VH

gene segment, 49.8, comprising an amino acid sequence identical to that contained in the published sequence of the hv1263 gene (Chen et al. (1989) Arthritis Rheum. 32: 72). Three of the VH3 genes, 184.3, 184.14, and 184.17, that were isolated with the o-184 probe contained sequences encoding identical amino acid sequences to those contained in the published for the VH genes DP-50, DP-54, and DP-45 (Tomlinson et al. (1992) J. Mol. Biol. 227: 776). These four VH genes were used to build the pH3V4D plasmid.

The 184.3 gene was found to be contained within a 3 kb BamHI fragment. This fragment was subcloned into the plasmid vector pGP1f such that the XhoI site of the polylinker is 5' of the gene, and the SalI site is 3'. The resulting plasmid is called p184.3.36f. The 184.14 gene was found to be contained within a 4.8 kb HindIII fragment. This fragment was subcloned into the plasmid vector pUC19 in an orientation such that the gene could be further isolated as a 3.5 kb fragment by XhoI/SalI digestion at a genomic XhoI site 0.7 kb upstream of the gene and a polylinker derived SalI site 3' of the gene. The resulting plasmid is called p184.14.1. The 184.17 gene was found to be contained within a 5.7 kb HindIII fragment. This fragment was subcloned into the plasmid vector pSP72 (Promega, Madison, WI) in an orientation such that the polylinker derived XhoI and SalI sites are, respectively, 5' and 3' of the gene. The insert of this plasmid includes an XhoI site at the 3' end of the gene which was eliminated by partial digestion with XhoI, Klenow fragment filling-in, and religation. The resulting plasmid is called p184.17SK. The 49.8 gene was found to be contained within 6.3 kb XbaI fragment. This fragment was subcloned into the plasmid vector pNNO3, such that the polylinker derived XhoI and ClaI sites are, respectively, 5' and 3' of the gene, to create the plasmid pVH49.8 (Taylor et al. (1994) International Immunol. 6: 579). The XhoI/ClaI insert of pVH49.8 was then subcloned into pGP1f to create the plasmid p49.8f, which includes unique XhoI and SalI sites respectively at the 5' and 3' end of the 49.8 gene.

The 3.5 kb XhoI/SalI fragment of p184.14.1 was cloned into the XhoI site of p184.3.36f to generate the plasmid

pRMVH1, which includes both the 184.14 and the 184.3 genes in the same orientation. This plasmid was digested with XhoI and the 5.7 kb XhoI/SalI fragment of p184.17SK was inserted to create the plasmid pRMVH2, which contains, from 5' to 3', the three VH genes 184.17, 184.14, and 184.3, all in the same orientation. The plasmid pRMVH2 was then cut with XhoI, and the 6.3 kb XhoI/SalI insert of p49.8f inserted to create the plasmid pH3VH4, which contains, from 5' to 3', the four VH genes 49.8, 184.17, 184.14, and 184.3, all in the same orientation.

The 10.6 kb XhoI/EcoRV insert of the human D region clone pDH1 (described *supra*; e.g., in Example 12) was cloned into XhoI/EcoRV digested pGPe plasmid vector to create the new plasmid pDH1e. This plasmid was then digested with EcoRV and ligated with a synthetic linker fragment containing a SalI site (5' - ccg gtc gac ccg - 3'; SEQ ID NO:185). The resulting plasmid, pDH1es, includes most of the human D1 cluster within an insert that can be excised with XhoI and SalI, such that the XhoI site is on the 5' end, and the SalI site is on the 3' end. This insert was isolated and cloned into the SalI site of pH3VH4 to create the plasmid pH3VH4D, which includes four germline configuration human VH gene segments and 8 germline configuration human D segments, all in the same 5' to 3' orientation. The insert of this clone can be isolated, substantially free of vector sequences, by digestion with NotI.

pCOR1xa

The plasmid pCOR1 (described *supra*) which contains a 32 kb XhoI insert that includes 9 human D segments, 6 human J segments, the J μ intronic heavy chain enhancer, the μ switch region, and the C μ coding exons-was partially digested with XhoI, Klenow treated, and a synthetic SalI linker ligated in to produce the new plasmid pCOR1xa, which has a unique XhoI site at the 5' end and a unique SalI site at the 3' end. Both pCOR1 and pCOR1xa contain a 0.6 kb rat heavy chain 3' enhancer fragment at the 3' end, which is included in the insert if the plasmid is digested with NotI instead of XhoI or XhoI/SalI.

pP1-570

A phage P1 library (Genome Systems Inc., St. Louis, Missouri) was screened by PCR using the oligonucleotide primer pair:

5' - tca caa gcc cag caa cac caa g - 3' (SEQ ID NO:186)

5' - aaa agc cag aag acc ctc tcc ctg - 3' (SEQ ID NO:187)

This primer pair was designed to generate a 216 bp PCR product with a human γ gene template. One of the P1 clones identified was found to contain both the human $\gamma 3$ and $\gamma 1$ genes within an 80 kb insert. The insert of this clone, which is depicted in Fig. 80, can be isolated, substantially free of vector sequences, by digestion with NotI and SalI.

p11-14

Restriction mapping of the human $\gamma 3/\gamma 1$ clone P1-570 revealed a 14 kb BamHI fragment near the 5' end of the insert. This 14 kb fragment was subcloned into the plasmid vector pGP1f such that the polylinker derived SalI site is adjacent to the 5' end of the insert. The resulting plasmid is called pB14. Separately, an 11 kb NdeI/SpeI genomic DNA fragment covering the 3' end of the human μ gene and the 5' end of the human δ gene, derived from the plasmid clone pJ1NA (Choi et al. (1993) Nature Genetics 4: 117), was subcloned into the SalI site of pBluescript (Stratagene, LaJolla, CA) using synthetic oligonucleotide adapters. The resulting SalI insert was then isolated and cloned into the SalI site of pB14 such that the relative 5' to 3' orientation of the μ fragment from pJ1NA is the same as that of the γ fragment from P1-570. The resulting clone is called p11-14. The insert of this clone can be isolated, substantially free of vector sequences, by digestion with NotI.

pHP-3a

The mouse heavy chain 3' enhancer (Dariavach et al. (1991) Eur. J. Immunol. 21: 1499; Liebersen et al. (1991) Nucleic Acids Res. 19: 933) was cloned from a balb/c mouse genomic DNA phage λ library. To obtain a probe, total balb/c

mouse thymus DNA was used as a template for PCR amplification using the following two oligonucleotides:

cck76: 5' - caa tag ggg tca tgg acc c -.3' (SEQ ID NO:188)

cck77: 5' - tca ttc tgt gca gag ttg gc - 3' (SEQ ID NO:189)

The resulting 220 bp amplification product was cloned using the TA Cloning™ Kit (Invitrogen, San Diego, CA) and the insert used to screen the mouse phage library. A positively hybridizing 5.8 kb HindIII fragment from one of the resultant phage clones was subcloned into pGP1f. The orientation of the insert of this subclone, pH3'ENfa, is such that the polylinker XhoI site is adjacent to the 5' end of the insert and the SalI site adjacent to the 3' end. Nucleotide sequence analysis of a portion of this HindIII fragment confirmed that it contained the 3' heavy chain enhancer. The insert of pH3'ENfa includes an XhoI site approximately 1.9 kb upstream of the EcoRI site at the core of the enhancer sequence. This XhoI site was eliminated by partial digestion, Klenow fill-in, and religation, to create the clone pH3'Efx, which includes unique XhoI and SalI sites, respectively, at the 5' and 3' ends of the insert.

The 3' end of the human γ 3/ γ 1 clone P1-570 was subcloned as follows: P1-570 DNA was digested with NotI, klenow treated, then digested with XhoI; and the 13 kb end fragment isolated and ligated to plasmid vector pGP2b which had been digested with BamHI, klenow treated, and then digested with XhoI. The resulting plasmid, pPX-3, has lost the polylinker NotI site adjacent to the polylinker XhoI site at the 5' end of the insert; however, the XhoI site remains intact, and the insert can be isolated by digestion with NotI and XhoI, or SalI and XhoI. The 3' enhancer containing XhoI/SalI insert of pH3'Efx was isolated and ligated into the 3' SalI site of pPX-3 to create the plasmid pHP-3a. The enhancer containing fragment within the pHP-3a insert is ligated in the opposite orientation as the 3' end of the P1-570 clone. Therefore, pHP-3a contains an internal SalI site, and the insert is isolated by digestion with XhoI and NotI.

Because this is an enhancer element, 5' to 3' orientation is generally not critical for function.

HCo7.

To prepare the HCo7 DNA mixture for pronuclear microinjection, DNA from each of the five plasmids described above was digested with restriction enzymes and separated on an agarose gel. Clone pH3V4D was cut with NotI; pCOR1xa was cut with NotI; p11-14 was cut with NotI; pP1-570 was cut with NotI and SalI; and pHP-3a was cut with NotI and XhoI. The DNA inserts were electroeluted and further purified on an equilibrium CsCl gradient without EtBr. The inserts were dialyzed into injection buffer and mixed as follows: 50 microliters of pH3V4D insert @ 20.4 ng/microliter; 50 microliters of pCOR1xa insert @ 20.8 ng/microliter; 50 microliters of p11-14 insert @ 15.6 ng/microliter; 300 microliters of pP1-570 insert @ 8.8 ng/microliter; 60 microliters of pHP-3a insert @ 10.8 ng/microliter; and 1.49 ml injection buffer.

HCo7 transgenic animals

The HCo7 DNA mixture was microinjected into the pronuclei of one-half day old embryos, and the embryos transferred into the oviducts of pseudopregnant females, as described by Hogan et al. (Manipulating the mouse embryo, Cold Spring Harbor laboratories, Cold Spring Harbor NY).

Tail tip DNA was isolated from 202 animals that developed from microinjected embryos. Southern blot analysis of this DNA, using a probe comprising human μ and DH sequences, revealed 22 founder animals that had incorporated at least a portion of the HCo7 transgene. Fig. 81 shows an analysis of the expression of human μ and human $\gamma 1$ in the serum of 6 G0 animals that developed from embryos microinjected with HCo7 DNA. Serum levels of human immunoglobulin proteins were measured by ELISA as described in Lonberg et al. (1994) Nature 368: 856. Four of these six mice showed evidence of incorporation of the transgene by Southern blot analysis, and three of these mice expressed both human μ and human $\gamma 1$

proteins in their serum. The single transgenic mouse that did not express human immunoglobulin proteins was determined by Southern blot analysis to contain only a low number of copies of the transgene, and it is possible that the entire transgene was not incorporated, or that this mouse was a genetic mosaic. Two of the founder HCo7 mice, #11952 and #11959, were bred with human κ minilocus (KCo4 line 4436) transgenic mice that were also homozygous for disruptions of the endogenous heavy, and κ light chain loci (Lonberg et al. *op.cit*), to generate mice that were homozygous for the two endogenous locus disruptions and hemizygous for the two introduced human miniloci, KCo4 and HCo7. Five of these so-called double-transgenic/double-deletion mice were analyzed for expression of human IgM, human IgG1, and human IgG3. As a control, three HC2/KCo4 double-transgenic/double-deletion mice were included in the analysis. This experiment is presented in Fig. 82. The ELISA data in this figure was collected as in Lonberg et al. (*op.cit*), except that for detection of human IgG3, the coating antibody was a specific mAb directed against human IgG3 (cat. # 08041, Pharmingen, La Jolla, CA); the other details of the IgG3 assay were identical to those published for IgG1. While the HC2/KCo4 mice express only human IgM and human IgG1, the HCo7/KCo4 mice also express human IgG3 in addition to these two isotypes. Expression of human γ 3 and γ 1 in the HCo7 mice has also been detected by PCR amplification of cDNA synthesized from RNA isolated from the spleen of a transgenic mouse. Fig. 83 depicts PCR amplification products synthesized using spleen cDNA from three different lines of transgenic mice: line 2550 is an HC2 transgenic line, while lines 11959 and 11952 are HCo7 transgenic lines. Single stranded cDNA was synthesized from spleen RNA as described by Taylor et al. (1992) Nucleic Acid Res. 20: 6287. The cDNA was then PCR amplified using the following two oligonucleotides:

o-382: 5'- gtc cag aat tcg gt(c,g,t) cag ctg gtg (c,g)ag tct. gg -.3' (SEQ ID NO:190)

o-383: 5'- ggt ttc tcg agg aag agg aag act gac ggt cc -
3' (SEQ ID NO:190)

This primer pair directs the synthesis of PCR products that spans the hinge region of human γ transcripts. Because of differences in the structures of the human $\gamma 1$ and $\gamma 3$ hinge regions, PCR amplification distinguishes between these two transcripts. A human $\gamma 1$ template will direct the synthesis of a 752 bp PCR product, while human $\gamma 3$ directs the synthesis of a 893 bp product. While only human $\gamma 1$ template is detectable in the HC2 line 2550 and HCo7 line 11959 spleens, both $\gamma 1$ and $\gamma 3$ transcripts are detectable in the HCo7 line 11952 spleen. Because of the non-quantitative nature of this assay, and because of differences in $\gamma 3$ expression between individual animals (shown by ELISA in Fig. 82), the inability to observe $\gamma 3$ in the HCo7 line 11959 spleen in Fig. 83 does not indicate that $\gamma 3$ is not expressed in this line. Isolated spleen cells from the HCo7/KCo4 mice can also be induced to express both IgG1 and IgG3 in vitro by stimulation with LPS and IL4. This experiment is shown in Fig. 84. Spleen cells from a 7 week old male HCo7/KCo4 double-transgenic/double-deletion mouse (#12496; line 11959/4436) tested for immunoglobulin secretion in response to the thymus-independent B cell mitogen, LPS, alone and in conjunction with various cytokines. Splenocytes were enriched for B cells by cytotoxic elimination of T cells. B-enriched cells were plated in 24 well plates at 2×10^6 cells per well in 2 ml of 10% FCS in RPMI-1640. LPS was added to all wells at 10 micrograms/ml. IL-2 was added at 50 units/ml, IL-4 was added at 15 ng/ml, IL-6 was added at 15 ng/ml, γ IFN was added at 100 units/ml. Cultures were incubated at 37°C, 5% CO₂ for 10 days, then supernatants were analyzed for human IgG1 and IgG3 by ELISA. All reagents for ELISA were polyclonal anti-serum from Jackson Immunologicals (West Grove, PA), except the capture anti-human IgM, which was a monoclonal antibody from The Binding Site (Birmingham, UK).

EXAMPLE 38

This example demonstrates the successful introduction into the mouse genome of functional human light chain V segments by co-injection of a human κ light chain minilocus and a YAC clone comprising multiple human V_{κ} segments. The example shows that the V_{κ} segment genes contained on the YAC contribute to the expressed repertoire of human κ chains in the resultant mouse. The example demonstrates a method for repertoire expansion of transgene-encoded human immunoglobulin proteins, and specifically shows how a human κ chain variable region repertoire can be expanded by co-introduction of unlinked polynucleotides comprising human immunoglobulin variable region segments.

Introduction of functional human light chain V segments by co-injection of V_{κ} containing yeast artificial chromosome clone DNA and κ light chain minilocus clone DNA

I. Analysis of a yeast strain containing cloned human V_{κ} gene segments.

Total genomic DNA was isolated from a yeast strain containing a 450 kb yeast artificial chromosome (YAC) comprising a portion of the human V_{κ} locus (ICRF YAC library designation 4x17E1). To determine the identity of some of the V_{κ} gene segments included in this YAC clone, the genomic DNA was used as a substrate for a series of V_{κ} family specific PCR amplification reactions. Four different 5' primers were each paired with a single consensus 3' primer in four sets of amplifications. The 5' primers were: o-270 (5'-gac atc cag ctg acc cag tct cc-3'; SEQ ID NO:192), o-271 (5'-gat att cag ctg act cag tct cc-3'; SEQ ID NO:193), o-272 (5'-gaa att cag ctg acg cag tct cc-3'; SEQ ID NO:194), and o-273 (5'-gaa acg cag ctg acg cag tct cc-3'; SEQ ID NO:195). These primers are used by Marks et al. (Eur. J. Immunol. 1991. 21, 985) as V_{κ} family specific primers. The 3' primer, o-274 (5'-gca agc ttc tgt ccc aga ccc act gcc act gaa cc-3'; SEQ ID NO:196), is based on a consensus sequence for FR3. Each of the four sets of primers directed the amplification of the expected 0.2 kb fragment from yeast genomic DNA containing the YAC clone 4x17E1. The 4 different sets of amplification products were then gel purified

and cloned into the PvuII/HindIII site of the plasmid vector pSP72 (Promega). Nucleotide sequence analysis of 11 resulting clones identified seven distinct V genes. These results are presented below in Table 14.

Table 14. Identification of human V_k segments on the YAC 4x17E1.

PCR primers	clone #	identified gene	V _k family
o-270/o-274	1	L22*	I
- "-	4	L22*	I
- "-	7	O2* or O12	I
o-271/o-274	11	A10*	VI
- "-	15	A10*	VI
o-272/o-274	20	A4* or A20	I
- "-	21	A11*	III
- "-	22	A11*	III
- "-	23	A11*	III
- "-	25	O4* or O14	I
o-273/o-274	36	L16* or L2	III

* Gene segments mapped within the distal V_k cluster (Cox et al. Eur. J. Immunol. 1994. 24, 827; Pargent et al. Eur. J. Immunol. 1991. 21, 1829; Schable and Zachau Biol. Chem. Hoppe-Seyler 1993. 374, 1001)

All of the sequences amplified from the YAC clone are either unambiguously assigned to V_k genes that are mapped to the distal cluster, or they are compatible with distal gene sequences. As none of the sequences could be unambiguously assigned to proximal V genes, it appears that the YAC 4x17E1 includes sequences from the distal V_k region. Furthermore, one of the identified sequences, clone #7 (VkO2), maps near the J proximal end of the distal cluster, while another sequence, clones # 1 and 4 (VkL22), maps over 300 kb upstream, near the J distal end of the distal cluster. Thus, if the 450 kb YAC

clone 4x17E1 represents a non-deleted copy of the corresponding human genome fragment, it comprises at least 32 different V_k segments. However, some of these are non-functional pseudogenes.

2. Generation of transgenic mice containing YAC derived V_k gene segments.

To obtain purified YAC DNA for microinjection into embryo pronuclei, total genomic DNA was size fractionated on agarose gels. The yeast cells containing YAC 4x17E1 were imbedded in agarose prior to lysis, and YAC DNA was separated from yeast chromosomal DNA by standard pulse field gel electrophoresis (per manufacturers specifications: CHEF DR-II electrophoresis cell, BIO-RAD Laboratories, Richmond CA). Six individual pulse field gels were stained with ethidium bromide and the YAC clone containing gel material was cut away from the rest of the gel. The YAC containing gel slices were then imbedded in a new (low melting temperature) agarose gel cast in a triangular gel tray. The resulting triangular gel was extended at the apex with a narrow gel containing two moles/liter sodium acetate in addition to the standard gel buffer (Fig. 85).

The gel was then placed in an electrophoresis chamber immersed in standard gel buffer. The "Y"-shaped gel former rises above the surface of the buffer so that current can only flow to the narrow high salt gel slice. A Plexiglas block was placed over the high salt gel slice to prevent diffusion of the NaOAc into the gel buffer. The YAC DNA was then electrophoresed out of the original gel slices and into the narrow high salt block. At the point of transition from the low salt gel to the high salt gel, there is a resistance drop that effectively halts the migration of the YAC DNA through the gel. This leads to a concentration of the YAC DNA at the apex of the triangular gel. Following electrophoresis and staining, the concentrated YAC DNA was cut away from the rest of the DNA and the agarose digested with GELase (EPICENTRE Technologies). Cesium chloride was then added to the YAC DNA containing liquid to obtain a density of 1.68 g/ml. This solution was

centrifuged at 37,000 rpm for 36 hrs to separate the DNA from contaminating material. 0.5 ml fractions of the resulting density gradient were isolated and the peak DNA containing fraction dialyzed against 5 mM tris (pH 7.4)/5 mM NaCl/0.1 M EDTA. Following dialysis, the concentration of the resulting 0.65 ml solution of YAC DNA was found to be 2 micrograms/ml. This DNA was mixed with purified DNA insert from plasmids pKC1B and pKV4 (Lonberg et al. 1994. Nature 368, 856) at a ratio of 20: 1: 1 (micrograms YAC4x17E1: KC1B: KV4). The resulting 2 microgram/ml solution was injected into the pronuclei of half-day mouse embryos, and 95 surviving microinjected embryos transferred into the oviducts of pseudo-pregnant females. Thirty nine mice were born that developed from the microinjected embryos. Two of these mice, #9269 and #9272, were used to establish transgenic lines. The lines are designated KCo5-9269 and KCo5-9272.

A Southern blot analysis of genomic DNA from mice of lines KCo5-9269 and KCo5-9272 was carried out to determine if YAC 4x17E1 derived V_k segments had been incorporated in their genomes. A V_k gene segment, V_kA10 (accession #: x12683; Straubinger et al. 1988. Biol. Chem. Hoppe-Seyler 369, 601-607), from the middle of the distal V_k cluster was chosen as a probe for the Southern blot analysis. To obtain the cloned probe, the V_kA10 gene was first amplified by PCR. The two oligo nucleotides, o-337 (5' - cgg tta aca tag ccc tgg gac gag ac - 3'; SEQ ID NO:197) and o-338 (5' - ggg tta act cat tgc ctg caa agc ac - 3'; SEQ ID NO:198), were used as primers to amplify a 1 kb fragment from YAC 4x17E1. The amplification product was gel purified, digested with HincII, and cloned into pUC18 to obtain the plasmid p17E1A10. The insert of this plasmid was then used to probe a southern blot of KCo5-9269 and KCo5-9272 DNA. The blot showed hybridization of the probe to the expected restriction fragments in the KCo5-9272 mouse DNA only. This indicates that the V_kA10 gene is incorporated into the genome of KCo5-9272 mice and not KCo5-9269 mice. Line KCo5-9272 mice were then bred with HC2-2550/JHD/JKD mice to obtain mice homozygous for disruptions of the endogenous heavy and κ light chain loci, and hemi- or homozygous for the HC2 and

KCo5 transgenes. Animals that are homozygous for disruptions of the endogenous heavy and k light chain loci, and hemi- or homozygous for human heavy and k light chain transgenes are designated double transgenic/double deletion mice.

A cDNA cloning experiment was carried out to determine if any of the YAC-derived V_k genes are expressed in line KCo5-9272 mice. The double transgenic/double deletion mouse #12648 (HC2-2550/KCo5-9272/JHD/JKD) was sacrificed and total RNA isolated from the spleen. Single stranded cDNA was synthesized from the RNA and used as a template in four separate PCR reactions using oligonucleotides o-270, o-271, o-272, and o-273 as 5' primers, and the Ck specific oligonucleotide, o-186 (5' - tag aag gaa ttc agc agg cac aca aca gag gca gtt cca - 3'; SEQ ID NO:173), as a 3' primer. The amplification products were cloned into the pCRII TA cloning vector (Invitrogen). The nucleotide sequence of 19 inserts was determined. The results of the sequence analysis are summarized in Table 15 below.

Table 15. Identification of human Vk genes expressed in mouse line KCo5-9272.

PCR primers	clone #	identified gene	Vk family
o-270/o-186	1	L15*	I
- "-	3	L18**	I
- "-	7	L24**	I
- "-	9	L15*	I
- "-	10	L15*	I
o-271/o-186	15	A10**	VI
- "-	17	A10**	VI
- "-	18	A10**	VI
- "-	19	A10**	VI
- "-	21	A10**	VI
o-272/o-186	101	A27*	III
- "-	102	L15*	I
- "-	103	A27*	III
- "-	104	A27*	III
o-273/o-186	35	A27*	III
- "-	38	A27*	III
- "-	44	A27*	III
- "-	45	A27*	III
- "-	48	A27*	III

* Vk genes encoded by transgene plasmid sequences.

** Vk genes encoded uniquely by YAC derived transgene sequences.

These results show that at least 3 of the YAC derived V_k gene segments, A10, L18, and L24, contribute to the expressed human repertoire of the line KCo5-9272 mice.

To determine the effect of this increased repertoire on the size of the various B220⁺ cell populations in the bone marrow and spleen, a flow

cytometric analysis was carried out on line KCo5-9272 mice. Part of this analysis is shown in Figs. 86 and 87. Two double transgenic/double deletion mice, one containing the KCo5 transgene, and one containing the KCo4 transgene, are compared in this experiment. These two transgenes share the same joining and constant region sequences, as well as the same intronic and 3' enhancer sequences. They also share four different cloned V gene segments; however, the KCo5 transgene includes the additional V segments derived from YAC 4x17E1 that are not included in the KCo4 transgene. Cells were isolated from mouse #13534 (HC2-2550/KCo5-9272/JHD/JKD) and mouse #13449 (HC2-2550/KCo4-4436/JHD/JKD). Bone marrow cells were stained with anti-mouse B220 (Caltag, South San Francisco, CA), anti-mouse CD43 (PharMingen, La Jolla, CA), and anti-human IgM (Jackson Immunologic, West Grove, PA). Spleen cells were stained with anti-mouse B220 and anti-human IgM.

Fig. 86 shows a comparison of the B cell, and B cell progenitor populations in the bone marrow of KCo5 and KCo4 mice. The fraction of B cells in the bone marrow (B220⁺, IgM⁺) is approximately three times higher in the KCo5 mice (6%) than it is in the KCo4 mice (2%). The pre-B cell population (B220⁺, CD43⁻, IgM⁻) is also higher in the KCo5 mice (9%, compared to 5% for KCo4). Furthermore, the pro-B compartment (B220⁺, CD43⁺) is elevated in these mice (11% for KCo5 and 5% for KCo4). Although each of these three compartments is larger in the KCo5 mice than it is in the KCo4 mice, the levels are still approximately half that found in wild type mice. The increase in the number of bone marrow B cells is presumably a direct consequence of the increased repertoire size. The larger primary repertoire of these mice may provide for membrane Ig with some minimal threshold affinity for endogenous antigens. Receptor ligation could then allow for proliferation of those B cells expressing the reactive Ig. However, because the pre-B and pro-B cells do not express light chain genes,

the explanation for the increased sizes of these two compartments in the KCo5 mice is not immediately apparent. The B cell progenitor compartments may be larger in KCo5 mice because the increased number of B cells creates a bone marrow environment that is more conducive to the expansion of these populations. This effect could be mediated directly by secreted factors or by cell-cell contact between B cells and progenitor cells, or it could be mediated indirectly, by titration of factors or cells that would otherwise inhibit the survival or proliferation of the progenitor cells.

Fig. 87 shows a comparison of the splenic B cell (B220⁺, IgM⁺) populations in KCo5 and KCo4 mice. The major difference between these two mice is the relative sizes of B220^{dull} B cell populations (6% in the KCo5 mice and 13% in the KCo4 mice). The B220^{dull} cells are larger than the B220^{bright} B cells, and a higher fraction of them express the λ light chain. These are characteristics of the so-called B1 population that normally dominates the peritoneal B cell population in wild type mice. The spleens of the KCo4 mice comprise an anomalously high fraction of B220^{dull} cells, while the KCo5 mice have a more normal distribution these cells. However, both strains contain approximately one-half to one-third the normal number of B cells in the spleen.

EXAMPLE 39

This example demonstrates the successful use of KCo5 transgenic mice of Example 38 to isolate hybridoma clones that secrete high affinity, antigen specific, human IgG monoclonal antibodies.

Immunization. A double deletion/double transgenic mouse (KCo5-9272/Hc2-2550/JHD/JKD, #12657) was immunized intraperitoneally every other week for eight weeks with 4 to 10 x 10⁶ irradiated T4D3 cells, a murine T cell line expressing human CD4 (Dr. Jane Parnes, Stanford

University) followed by one injection intraperitoneally two weeks later of 20 mg soluble recombinant human CD4 (sCD4; Intracell) in incomplete Freund's adjuvant (Sigma). The mouse was boosted once 3 days prior to fusion with 20 mg sCD4 intravenously.

Hybridoma fusion.

Single cell suspensions of splenic lymphocytes from the immunized mouse were fused to one-sixth the number of P3X63-Ag8.653 nonsecreting mouse myeloma cells (ATCC CRL 1580) with 50% PEG (Sigma). Cells were plated at approximately 2×10^5 in flat bottom microtiter plates, followed by a two week incubation in selective medium containing 20% Fetal Clone Serum (HyClone), 18% "653" conditioned medium, 5% Origen (IGEN), 4 mM L-glutamine, 1 mM sodium pyruvate, 5 mM HEPES, 0.055 mM 2-mercaptoethanol, 50 units/ml mM penicillin, 50 mg/ml streptomycin, 50 mg/ml mM gentamycin and 1X HAT (Sigma; the HAT was added 24 hrs after the fusion). After two weeks, cells were cultured in medium in which the HAT was replaced with HT. Wells were screened by ELISA and flow cytometry once extensive hybridoma growth or spent medium was observed.

Hybridoma screening by ELISA.

To detect anti-CD4 mAbs, microtiter plates (Falcon) were coated overnight at 4°C with 50 ml of 2.5 mg/ml of sCD4 in PBS, blocked at RT for 1 hr with 100 ml of 5% chicken serum in PBS, and then sequentially incubated at RT for 1 hr each with 1:4 dilutions of supernatant from hybridomas, 1:1000 dilution of F(ab')₂ fragments of horseradish peroxidase (HRPO)-conjugated goat anti-human IgG (Jackson) or 1:250 dilution of HRPO-conjugated goat anti-human Igk antibodies (Sigma) plus 1% normal mouse serum, and finally with 0.22 mg/ml ABTS in 0.1 M citrate phosphate buffer, pH 4 with 0.0024% H₂O₂. Plates were washed 3-6 times with wash buffer (0.5% Tween-20 in PBS) between all incubations, except the first. Diluent (wash buffer with 5% chicken serum) was used to dilute the supernatants

and the HRPPO conjugates. Absorbance was measured using dual wavelengths (OD at the reference wavelength of 490 nm was subtracted from the OD at 415 nm).

To detect mouse λ -containing mAbs, the above ELISA protocol was used, with the following exceptions. Wells of microtiter plates were coated with 100 μ l of 1) 1.25 mg/ml goat anti-mouse λ (Pierce), 2) 1.25 mg/ml goat anti-human Fc γ (Jackson), or 3) 2.5 mg/ml sCD4 (ABT). For the detection step, 100 μ l of 1:5000 goat anti-mouse λ (SBA) conjugated to biotin was used followed by 100 μ l of 1:1000 streptavidin conjugated to HRPPO (Jackson). Murine and human mAb standards were used at the indicated concentrations. To look for cross-reactivity to unrelated antigens, wells were coated with CEA (Crystal Chem), KLH (CalBiochem), HSA (Sigma), BSA (Sigma) or OVA (Sigma; all at 2 mg/ml, except CEA which was at 2.5). Appropriate antibodies were titrated and used as positive controls (human IgM anti-CEA (GenPharm), rabbit anti-KLH (Sigma), sheep anti-HSA (The Binding Site), sheep anti-BSA (The Binding Site), and sheep anti-OVA (The Binding Site)). Any bound antibody was detected with HRPPO conjugates of goat anti-human IgM, donkey anti-rabbit IgG or donkey anti-sheep IgG (all diluted 1:1000 and obtained from Jackson). Otherwise, the standard ELISA protocol was followed.

Hybridoma screening by flow cytometric assay.

To further screen for mAbs reactive with native cell-surface CD4, 5×10^5 SupT1 cells (ATCC CRL 1942) were incubated on ice with a 1:2 dilution of spent supernatant from the fusion plates for 30 min, washed twice with cold stain buffer (0.1% BSA, 0.02% NaN₃ in PBS), incubated with 1.5 mg/ml of an F(ab')₂ fragment of FITC-conjugated goat anti-human Fc γ (FITC-GaHuIgG; Jackson) for 15 min, washed once and analyzed immediately on a FACScan (Becton-Dickinson).

CD4 reactive hybridomas.

Using the ELISA and flow cytometric techniques described above, 12 hybridoma clones were identified that secreted human IgG specifically reactive with native human CD4. Ten of these twelve clones were further subcloned. Eight of these subclones were identified as human IgG1k secreting hybridomas. The other two expressed a mouse λ light chain. The parent wells for the 8 fully human clones were: 1E11, 2E4, 4D1, 6C1, 6G5, 7G2, 10C5, and 1G1. Flow cytometric assays of the binding of 3 of the fully human IgGk subclones (4D1.4, 6G5.1, and 10C5.6) are shown in Fig. 88.

Fig. 88 shows binding of IgGk anti-nCD4 monoclonal antibodies to CD4+ SupT1 cells. Cells from log phase growth cultures were washed and stained with no monoclonal antibody, 4E4.2 (as a negative control), chimeric Leu3a (as a positive control), or with one of the 10 human IgG anti-nCD4 monoclonal antibodies. Any bound monoclonal antibody was detected with FITC-conjugated goat anti-human Fcy. All ten monoclonal antibodies bound to SupT1 cells, although data is shown here for only three of them.

Analysis of human antibody secretion by cloned hybridomas.

To compare the growth and secretion levels of mAbs, the subclones were put into replicate cultures in HT medium in 24 well plates at an initial density of 2×10^5 cells/ml. Each day for 7 days, one of the replicate cultures for each subclone was harvested and cell numbers, cell viability (by Trypan blue exclusion) and the amount of mAb in the supernatant (by a quantitative ELISA for total human γ) were determined. Table 16 shows data for antibody secretion by 7 of the hybridoma subclones.

Table 16. Secretion Levels For Human IgGk Anti-nCD4 Monoclonal Antibodies

Subclone	pg/cell	pg/cell/d
1E11.15	3.9	0.56
1G1.9	11	1.5
4D1.4	1.4	0.91

6C1.10	3.3	0.48
6G5.1	7.8	1.1
7G2.2	4.4	0.63
10C5.6	8.0	1.1

* $\text{pg/cell} = (\text{maximum amount of mAb}) / (\text{maximum number of viable cells})$
 $\text{pg/cell/d} = (\text{pg/cell}) / 7 \text{ days}$

Purification of human mAbs.

The individual hybridoma clones were grown in medium without HT and Origen and the FCS was gradually decreased to approximately 2-3% in the final 1 l cultures. Supernatants were harvested once the viability of the hybridomas fell below approximately 30%. To purify the IgGk mAbs, the spent supernatants were centrifuged to remove cells, concentrated via ultrafiltration to approximately 50 to 100 mls, diluted 1:5 with PBS, pH 7.4 and loaded onto a 5 ml Protein A (Pharmacia) column. After washing with 3-5 column volumes of PBS, the human IgGk mAbs were eluted with 0.1 HCl, 150 mM NaCl, pH 2.8 and immediately neutralized with 1M Tris base. Column fractions containing material with an $\text{OD}_{280} > 0.2$ were pooled and dialyzed into PBS. The OD_{280} was then determined and an absorbtivity coefficient of 1.4 was used to calculate the protein concentration of the human IgG. No mAb was detected in the flow through and the % recoveries ranged from 93 to 100%. Three to six mgs of each purified mAb were obtained, with >90% purity.

Analysis of monoclonal antibodies from cloned hybridomas.

To investigate the specificity of binding of mAbs, human PBMC were isolated over Ficoll and stained as follows. Human PBMC (10^6) in stain buffer were incubated for 30 min on ice, in separate reactions, with equal volumes of supernatant from each of three of the subcloned hybridomas (4D1.4, 6G5.1, and 10C5.6), or with an isotype matched negative control mAb, washed twice, and incubated 20 min on ice with 1 mg/ml of FITC-GaHuIgG along with either 10 ml of mouse anti-human CD4 mAb (Leu3a; Becton-Dickinson) conjugated to phycoerythrin (PE), 10 ml of mouse anti-human

CD8 mAb (Leu2a; Becton-Dickinson) conjugated to PE, or 5 ml of mouse anti-human CD19 mAb (SJ25-C1; Caltag) conjugated to PE. Gated lymphocytes were then analyzed on a FACScan flow cytometer (Becton Dickinson, San Jose, CA). All three of the antibodies were found to bind specifically to the CD4 fraction of the human PBMC.

To approximate the location of the epitope recognized by these three mAbs, 5×10^5 SupT1 cells were pre-incubated for 20 min on ice with buffer, 2.5 mg/ml RPA-T4, or 2.5 mg/ml Leu3a in stain buffer, then for 30 min with one of the 10 human IgG mAbs (in supernatant diluted 1:2) and finally with 0.5 mg/ml FITC-conjugated goat anti-human Fc γ to detect any bound human IgG. Cells were washed twice with stain buffer prior to and once after the last step. The results of this blocking assay are shown in Fig. 89. None of the three antibodies share an epitope with RPA-T4, while 6G5.1 and 10C5.6 appear to recognize the same (or an adjacent) epitope as that recognized by Leu3a.

Rate and equilibrium constant determinations.

Human sCD4 (2500 to 4200 RU) was immobilized by covalent coupling through amine groups to the sensor chip surface according to manufacturer's instructions. Antibody dilutions were flowed over the antigen-coupled sensor chips until equilibrium was reached, and then buffer only was allowed to flow. For each phase of the reaction, binding and dissociation, the fraction of bound antibody was plotted over time. The derivative of the binding curve (dR/dt) was calculated and plotted against the response for each concentration. To calculate the association rate constant (k_{assoc}), the slopes of those resulting lines were then plotted against the concentration of the monoclonal antibody. The slope of the line from this graph corresponded to the k_{assoc} . The dissociation rate constant (k_{dissoc}) was calculated from the log of the drop in response (during the buffer flow phase) against the time interval. The K_a was derived by dividing the k_{assoc} by the k_{dissoc} . The measured rate and affinity constant data for 5 different

purified monoclonal antibodies derived from the KCo5/HC2 double transgenic/double deletion mice, and one purified antibody obtained from a commercial source (Becton Dickinson, San Jose, CA), is presented in Table 17.

Table 17. Rate and affinity constants for monoclonal antibodies that bind to human CD4.

Hybridoma	Antibody	Source	k_{ass} ($M^{-1}s^{-1}$)	k_{diss} (s^{-1})	K_a (M^{-1})
1E11.15	human IgG1k	HC2/KCo5 transgenic	2.7×10^5	4.6×10^{-5}	5.8×10^9
1G1.9	human IgG1k	HC2/KCo5 transgenic	9.1×10^4	2.2×10^{-5}	4.2×10^9
4D1.4	human IgG1k	HC2/KCo5 transgenic	9.8×10^4	4.2×10^{-5}	2.3×10^9
6G5.1	human IgG1k	HC2/KCo5 transgenic	1.1×10^5	1.0×10^{-5}	1.1×10^{10}
10C5.6	human IgG1k	HC2/KCo5 transgenic	7.4×10^4	1.6×10^{-5}	4.5×10^9
Leu3a	mouse IgG1k	Becton Dickinson	1.5×10^5	4.2×10^{-6}	3.7×10^{10}

Mixed Lymphocyte Reaction (MLR).

To compare the *in vitro* efficacy of the human monoclonal antibody 10C5.6, derived from the KCo5 transgenic mouse, to that of the mouse antibody Leu3a, an MLR assay was performed. Human PBMC from 2 unrelated donors were isolated over Ficoll and CD4+ PBL from each donor were purified using a CD4 column (Human CD4 Collect, Biotex Laboratories, Inc., Canada) according to manufacturer's directions.

Inactivated stimulator cells were obtained by treating PBMC from both donors with 100 mg/ml mitomycin C (Aldrich) in culture medium (RPMI 1640 with 10% heat-inactivated human AB serum (from NABI), Hepes, sodium pyruvate, glutamine, pen/strep and b-mercaptoethanol (all used at manufacturer's recommended concentrations)) for 30 min at 37°C followed by 3 washes with culture medium. Varying concentrations of

mAbs diluted in culture medium or culture medium only were sterile filtered and added at 100 μ l per well in triplicate in a 96 well round bottom plate. Fifty μ l of 10^5 CD4+ PBL from one donor in culture medium and 10^5 mitomycin C-treated PBMC from the other donor in 50 μ l of culture medium were then added to each well. Control plates with CD4+ PBL responders alone plus mAbs were set up to control for any toxic or mitogenic effects of the mAbs. A stimulator only control and a media background control were also included. After seven days in a 37°C, 5% CO₂ humidified incubator, 100 μ l of supernatant from each well was removed and 20 μ l of colorimetric reagent (Cell Titer 96AQ kit, Promega Corporation, Madison, WI) was added. Color was allowed to develop for 4 to 6 hrs and plates were read at 490 nm. The results of this experiment, depicted in Fig. 90, show that the human IgG1 κ antibody 10C5.6 is at least as effective as Leu3a at blocking the function of human PBMC CD4 cells in this assay.

Example 40.

Binding Characteristics of Human IgGkappa Anti-CD4 monoclonal Antibodies.

This example provides the binding characteristics of human IgG κ monoclonal antibodies derived from hybridoma clones obtained from HC2/KCo5/JHD/JCKD transgenic mice immunized with human CD4. The monoclonal antibodies are shown to have high avidity and affinity for recombinant and natural human CD4.

Cells from 10 individual hybridoma cell lines (1E11, 1G2, 6G5, 10C5, 1G1, 6C1, 2E4, 7G2, 1F8 and 4D1) that secrete human IgG kappa monoclonal antibodies (mAB) reactive with human CD4, were derived from JHD/JCKD/HC2/KCo5 transgenic mice. The cell lines were grown in culture, and antibody proteins were isolated from the supernatant (Fishwild, et al. 1996, *Nature Biotechnology* 14, 845-851, which is incorporated herein by reference). Antibody purified by Protein A affinity chromatography was used to

measure binding constants. The results are displayed in Tables 18 and 19.

The rate and equilibrium constants presented in Table 18 were determined with a BIAcore (Pharmacia Biosensor) using goat anti-human IgG (Fc-specific) coupled to the sensor chip and flowing a saturating concentration of mAb over followed by various concentrations of antigen (rCD4). These constants were derived from three experiments using purified mAbs.

Table 18. Affinity and Rate Constants.

Rate Constants (mean \pm SD)

Human mAb	k_{assoc} ($\text{M}^{-1}\text{s}^{-1}$)	k_{dissoc} (s^{-1})	K_a (M^{-1})
1E11.15	$1.7 (\pm 0.15) \times 10^5$	$3.5 (\pm 0.09) \times 10^{-3}$	5.0×10^7
6C1.10	$1.8 (\pm 0.44) \times 10^5$	$3.3 (\pm 0.04) \times 10^{-3}$	5.4×10^7
1G1.9	$1.2 (\pm 0.18) \times 10^5$	$9.4 (\pm 0.22) \times 10^{-4}$	1.3×10^8
6G5.1	$9.3 (\pm 1.1) \times 10^4$	$6.9 (\pm 0.36) \times 10^{-4}$	1.4×10^8
10C5.6	$9.4 (\pm 0.98) \times 10^4$	$7.1 (\pm 0.36) \times 10^{-4}$	1.3×10^8
2E4.2	$1.8 (\pm 0.10) \times 10^5$	$2.5 (\pm 0.05) \times 10^{-3}$	7.1×10^7
4D1.4	$2.5 (\pm 0.55) \times 10^5$	$3.4 (\pm 0.15) \times 10^{-3}$	7.3×10^7
7G2.2	$2.4 (\pm 0.31) \times 10^5$	$3.3 (\pm 0.07) \times 10^{-3}$	7.3×10^7
1F8.3	$1.8 (\pm 0.24) \times 10^5$	$4.3 (\pm 0.14) \times 10^{-3}$	4.3×10^7
1G2.10	$2.2 (\pm 0.26) \times 10^5$	$2.3 (\pm 0.03) \times 10^{-3}$	9.8×10^7
chi Leu3a	$1.5 (\pm 0.35) \times 10^5$	$2.3 (\pm 0.12) \times 10^{-4}$	6.6×10^8

The rate and equilibrium constants presented in Table 19 were determined with a BIAcore, using antigen (rCD4) coupled to the sensor chip and flowing mAb over. These constants were derived from at least three independent experiments using purified mAbs.

Table 19. Avidity and Rate Constants

Rate Constants (mean \pm SD)				
Human mAb	k_{assoc} ($\text{M}^{-1}\text{s}^{-1}$)	k_{dissoc} (s^{-1})	K_a (M^{-1})	
1E11.15	$2.8 (\pm 0.22) \times 10^5$	$4.5 (\pm 0.43) \times 10^{-5}$	6.2×10^9	
6C1.10	$2.0 (\pm 0.25) \times 10^5$	$4.0 (\pm 0.63) \times 10^{-5}$	5.1×10^9	
1G1.9	$9.1 (\pm 0.95) \times 10^4$	$2.2 (\pm 0.71) \times 10^{-5}$	4.2×10^9	
6G5.1	$1.1 (\pm 0.41) \times 10^5$	$1.0 (\pm 0.34) \times 10^{-5}$	1.1×10^{10}	
10C5.6	$7.4 (\pm 1.5) \times 10^4$	$1.6 (\pm 0.57) \times 10^{-5}$	4.5×10^9	
2E4.2	$1.4 (\pm 0.15) \times 10^5$	$2.2 (\pm 0.25) \times 10^{-5}$	6.3×10^9	
4D1.4	$9.8 (\pm 0.69) \times 10^4$	$4.2 (\pm 1.3) \times 10^{-5}$	2.3×10^9	
7G2.2	$1.7 (\pm 0.20) \times 10^5$	$5.0 (\pm 0.42) \times 10^{-5}$	3.4×10^9	
1F8.2	$1.7 (\pm 0.13) \times 10^5$	$9.7 (\pm 1.2) \times 10^{-5}$	1.7×10^9	
1G2.10	$1.7 (\pm 0.04) \times 10^5$	$6.3 (\pm 0.49) \times 10^{-5}$	2.7×10^9	
chi Leu3a	$4.0 (\pm 0.45) \times 10^5$	$1.2 (\pm 0.25) \times 10^{-5}$	3.4×10^{10}	
Leu3a	$1.5 (\pm 0.30) \times 10^5$	$4.2 (\pm 0.49) \times 10^{-6}$	3.7×10^{10}	

Table 20 provides equilibrium constants for anti-CD4 mABs presented in the scientific literature.

Table 20. Avidity and Rate Constants Reported for Anti-CD4 monoclonal antibodies

anti-CD4 mAb	Rate Constants (mean \pm SD)		
	k_{assoc} ($\text{M}^{-1}\text{s}^{-1}$)	k_{dissoc} (s^{-1})	K_a (M^{-1})
CE9.1 ⁽⁴⁾	NR*	NR	3.1×10^{10}
cMT412 ⁽¹⁾	NR	NR	5.0×10^9
chi Leu3a ⁽²⁾	NR	NR	1.0×10^{11}
BL4 ⁽³⁾	NR	NR	5.5×10^7
BB14 ⁽³⁾	NR	NR	3.3×10^8
cA2 ⁽⁵⁾	NR	NR	1.8×10^9
CDP571 ⁽⁶⁾	NR	NR	7.1×10^9

* NR = not reported

- (1) J. Cell. Biol. 15E:A179.
 (2) J. Immunol. 145:2839.
 (3) Clin. Immunol. Immunopath. 64:248.
 (4) Biotechnology. 10:1455.
 (5) Mol. Immunol. 30:1443.
 (6) European Patent Appl. #0626389A1.

The avidity and affinity determinations described above were performed with recombinant CD4 (rCD4). To determine the avidity of the human monoclonal antibodies for native CD4 (nCD4). An additional binding assay was used that does not require the antibody to be modified. Specifically, serial dilutions of antibody were incubated with SupT1 cells for 6 hr on ice, washed and detected any bound antibody with FITC-goat anti-human Fcy. The K_a is determined from the concentration of antibody that gives one-half of the maximum fluorescence (a four parameter fit was used). The results demonstrate that all ten human monoclonal antibodies bind very well to nCD4, with K_a values $>10^9 \text{ M}^{-1}$ (Table 21). Most antibodies, including chimeric Leu3a, bound less well to nCD4

than to rCD4. This could be due to differences in antigen density as well as to differences between the two antigens.

Table 21. Avidity Constants Determined
by Flow Cytometry.

Human mAb	<u>Ka values (M⁻¹)</u>		Ratio of Ka (rCD4/nCD4)
	rCD4	nCD4*	
1E11.15	6.2×10^9	3.3×10^9	1.9
6C1.10	5.1×10^9	3.1×10^9	1.6
1G1.9	4.2×10^9	2.3×10^9	1.9
6G5.1	1.1×10^{10}	1.9×10^9	5.9
10C5.6	4.5×10^9	1.8×10^9	2.5
2E4.2	6.3×10^9	1.1×10^9	5.8
4D1.4	2.3×10^9	2.0×10^9	1.2
7G2.2	3.4×10^9	3.3×10^9	1.0
1F8.2	1.7×10^9	3.2×10^9	0.5
1G2.10	2.7×10^9	1.9×10^9	1.4
chi Leu3a	3.4×10^{10}	5.6×10^9	6.1

* Human monoclonal antibodies were incubated in serial dilutions with SupT1 cells for 6 hrs, washed twice and incubated with FITC-conjugated goat anti-human Fcy antisera, washed and fixed. The Ka was calculated from the concentration of antibody yielding one-half of the maximum fluorescence as determined from a four-parameter fit.

Example 41.

Identification of Nucleotide Sequences Encoding Human IgGkappa Anti-CD4 Antibodies.

This example demonstrates that each of the hybridomas tested produces only one functional heavy [~~ex~~] and one functional light chain RNA transcript, consistent with

proper functioning allelic exclusion. In addition, sequence analysis of heavy and light chain CDR segments indicates that somatic mutation of the immunoglobulin transgenes has taken place.

Cells from five individual hybridoma cell lines (1E11, 1G2, 6G5, 10C5, and 4D1) that secrete human IgG kappa monoclonal antibodies reactive with human CD4, and derived from JHD/JCKD/HC2/KCo5 transgenic mice, were used to isolate RNA encoding each of the individual antibodies (Fishwild et al. 1996, *Nature Biotechnology* 14, 845-851). The RNA was used as a substrate to synthesize cDNA, which was then used to amplify human Ig gamma and kappa transcript sequences by PCR using primers specific for human VH, Vkappa, Cgamma, and Ckappa (Taylor et al. 1992, *Nucleic Acids Res.* 20, 6287-6295; Larrick, J.W., et al. (1989), *Bio/Technology*. 7. 934-938; Marks, J.D., et al. (1991). *Eur. J. Immunol.* 21. 985-991; Taylor, et al., 1994, *Int. Immunol.* 6, 579-591). The amplified Ig heavy and kappa light chain sequences were cloned into bacterial plasmids and nucleotide sequences determined. Analysis of the sequences spanning the heavy chain VDJ and light chain VJ junctions revealed in-frame heavy and light chain transcripts for each of the 5 clones, and in some cases additional out-of-frame sterile transcripts representing non-functional alleles. Consistent with proper functioning allelic exclusion, in no case was there more than one unique functional heavy or light chain transcript identified for each of the individual clones. Partial nucleotide sequences for each of the ten functional transcripts are assigned the following SEQ ID NOS: sequence I.D. No's: 1E11 gamma (SEQ ID NO:199); 1E11 kappa (SEQ ID NO:200); 1G2 gamma (SEQ ID NO:201); 1G2 kappa (SEQ ID NO:202); 6G5 gamma (SEQ ID NO:203); 6G5 kappa (SEQ ID NO:204); 10C5 gamma (SEQ ID NO:205); 10C5 kappa (SEQ ID NO:206); 4D1 gamma (SEQ ID NO:207); 4D1 kappa (SEQ ID NO:208) and are presented in Table 22. All sequences are presented in a 5' to 3' orientation.

Table 22. Partial Nucleotide Sequence for
Functional Transcripts

1E11 gamma (SEQ ID NO:199)

TGCACAAGAACATGAAACACCTGTGGTTCTTCCTCCTCCTGGTGGCAGCTCCCAGATGGGTC
CTGTCCCAGGTGCAGCTTCATCAGTGGGGCGCAGGACTGTTGAAGCCTTCGGAGACCCTGTC
CCTCACCTGCGCTGTCTATGGTGGGTCCTTCAGTGGTTACTTCTGGAGCTGGATCCGCCAGC
CCCCAGGGAGGGGGCTGGAGTGGATTGGGGAAATCCATCATCGTGGAAGCACCAACTACAAC
CCGTCCCTCGAGAGTCGAGTCACCCTATCAGTAGACACGTCCAAAACAGTTCTCCCTGAG
GCTGAGTTCTGTGACCGCCGCGGACACGGCTGTGTATTACTGTGCGAGAGACATTACTATGG
TTCGGGGAGTACCTCACTGGGGCCAGGGAACCCTGGTCACC

1E11 kappa (SEQ ID NO:200)

GACAGACTTCACTCTCACCATCAGCAGACTGGAGCCTGAAGATTTTGCAGTGTATTACTGTC
AGCAGTATGGTAGCTCACCCCTCACTTTCGGCGGAGGGACCAAGGTGGAGATCAAACGAACT
GTGGCGGCACCATCTGTCTTCATCTTCCC

1G2 gamma (SEQ ID NO:201)

TCCACCATCATGGGGTCAACCGCCATCCTCGCCCTCCTCCTGGCTGTTCTCCAAGGAGTCTG
TGCCGAGGTGCAGCTGGTGCAGTCTGGAGCAGAGGTGAAAAAGCCCGGGGAGTCTCTGAAGA
TCTCCTGTAAGGGTTCTGGATACAGCTTTACCAGTTACTGGATCGCCTGGGTGCGCCAGATG
CCCGGGAAAGGCCTGGAGTGGATGGGGATCATCGATCCTGCTGACTCTGATACCAGATACAA
CCCGTCCTTCCAAGGCCAGGTCACCATCTCAGCCGACAAGTCCATCAGTACCGCCTATTTGC
AGTGGAGCAGCCTGAAGGCCTCGGACACCGCCATGTATTACTGTGCGAGACCAGCGAACTGG
AACTGGTACTTCGTTCTCTGGGGCCGTGGCACCCCTGGTCACT

1G2 kappa (SEQ ID NO:202)

GACAGATTTCACTCTCACCATCAGCAGCCTGCAGCCTGAAGATTTTGCAACTTATTACTGTC
AACAGTTTATTAGTTACCCTCAGCTCACTTTCGGCGGAGGGACCAGGGTGGAGATCAAACGA
ACTGTGGCTGCACCATCTGTCTTCATCTTCCC

6G5 gamma (SEQ ID NO:203)

TGCACAAGAACATGAAACACCTGTGGTTCTTCCTCCTCCTGGTGGCAGCTCCCAGATGGGTC
CTGTCCCAGGTGCAGCTACAGCAGTGGGGCGCAGGACTGTTGAAGCCTTCGGAGACCCTGTC
CCTCACCTGCGCTGTCTATGGTGGGTCCTTCAGTGGTTACTACTGGAGCTGGATCCGCCAGC
CCCCAGGTAAGGGGCTGGAGTGGATTGGGGAAATCAATCATAGTGGAAGCACCAACTACAAC
CCGTCCCTCAAGAGTCGAGTCACCATATCAGTCGACACGTCCAAGAACCAGTTCTCCCTGAA
ACTGAGCTCTGTGACCGCCGCGGACACGGCTGTGTATTACTGTGCGAGAGTAATTAATTGGT
TCGACCCCTGGGGCCAGGGAACCCTGGTCACC

6G5 kappa (SEQ ID NO:204)

GACAGATTTCACTCTCACCATCAGCAGCCTGCAGCCTGAAGATTTTGCAACTTACTATTGTC
AACAGGCTAATAGTTTCCCGTACACTTTTGGCCAGGGGACCAAGCTGGAGATCAAACGAACT
GTGGCTGCACCATCTGTCTTCATCTTCCC

10C5 gamma (SEQ ID NO:205)

ATGAAACACCTGTGGTTCTTCCTCCTCCTGGTGGCAGCTCCCAGATGGGTCTGTCCCAGGT
GCAGCTACAGCAGTGGGGCGCAGGACTGTTGAAGCCTTCGGAGACCCTGTCCCTCACCTGCG
CTGTCTATGGTGGGTCTTCAGTGGTTACTACTGGAGCTGGATCCGCCAGCCCCCAGGTAAG
GGGCTGGAGTGGATTGGGGAAATCAATCATAGTGGAAGCACCAACTACAACCCGTCCCTCAA
GAGTCGAGTCACCATATCAGTCGACACGTCCAAGAACCAGTTCTCCCTGAAGCTGAGCTCTG
TGACCGCCGCGGACACGGCTGTGTATTACTGTGCGAGAGTAATTAATTGGTTTCGACCCCTGG
GGCCAGGGAACCCTGGTCACCGTCTCCTCAG

10C5 kappa (SEQ ID NO:206)

ATGGACATGATGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGTTCCCAGGTTCCAG
ATGCGACATCCAGATGACCCAGTCTCCATCTTCCGTGTCTGCATCTGTAGGAGACAGAGTCA
CCATCACTTGTCTGGGCGAGTCAGGATATTAGCAGCTGGTTAGCCTGGTATCAGCATAAACCA
GGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGGTCCCATCAAG
GTTTCAGCGGCAGTGGATCTGGGACAGATTTCACTCTCACCATCAGCAGCCTGCAGCCTGAAG
ATTTTGCAACTTACTATTGTCAACAGGCTAATAGTTTCCCGTACACTTTTGGCCAGGGGACC
AAGCTGGAGATCAAAC

4D1 gamma (SEQ ID NO:207)

ATGGGGTCAACCGCCATCCTCGCCCTCCTCCTGGCTGTTCTCCAAGGAGTCTGTGCCGAGGT
GCAGCTGGTGCAGTCTGGAGCAGAGGTGAAAAAGCCCGGGGAGTCTCTGAAGATCTCCTGTA
AGGGTTCTGGATACAGCTTTACCGGCTACTGGATCGGCTGGGTGCGCCAGATGCCCCGGGAAA
GGCCTGGAGTGGATGGGGATCATCTATCCTGGTGACTCTGATACCACATACAGCCCGTCTTT
CCAAGGCCAGGTCACCATCTCAGCCGACAAGTCCATCAGCACCGCCTACCTGCAGTGGAGCA
GCCTGAAGGCCTCGGACACCGCCATGTATTACTGTGCGAGAGACCAACTGGGCCTCTTTGAC
TACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGCCTCCACCAAGGGCCCATCGGTCTT
CCCCCTGGCACCTCCTCCAAGAAGCTT

4D1 kappa (SEQ ID NO:208)

ATGGACATGGAGTTCCCCGTTCACTCCTGGGGCTCCTGCTGCTCTGTTTCCCAGGTGCCAG
ATGTGACATCCAGATGACCCAGTCTCCATCCTCACTGTCTGCATCTGTAGGAGACAGAGTCA
CCATCACTTGTCTGGGCGAGTCAGGGTATTAGCAGCTGGTTAGCCTGGTATCAGCAGAAACCA
GAGAAAGCCCCTAAGTCCCTGATCTATTCTGCATCCAGTTTGCAAAGTGGGGTCCCATCAAG

GTTCAGCGGCAGTGGATCTGGGACAGATTTCACTCTCACCATCAGCAGCCTGCAGCCTGAAG
 ATTTTGCAACTTATTACTGCCAACAGTATGATAGTTACCCGTACACTTTTGGCCAGGGGACC
 AAGCTGGAGATCAAACGAACTGTGGCTGCACCATCTGTCTTCATCTTCCCGCCATCTGATGA
 AGCTT

Analysis of these DNA sequences demonstrates that the 5 hybridoma clones represent descendants of 4 individual primary B cells. Table 23 shows the amino acid sequences derived for each of the ten CDR3 regions, and the assignments for germline gene segments incorporated into each of the genes encoding these transcripts. The germline assignments are based on published gene sequences available from the National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Md. Also see: Cook et al. 1994, Nature Genet. 7, 162-168; Tomlinson et al. 1992, J. Mol. Biol. 227, 776-798; Matsuda et al. 1993, Nature Genet. 3, 88-94; Schable and Zachau, 1993, Biol. Chem. Hoppe-Seyler 374, 1001-1022; Cox et al. 1994, Eur. J. Immunol. 24, 827-836; Ravetch et al. 1981, Cell 27, 583-591; Ichihara et al. 1988, EMBO J. 7, 4141-4150; Yamada et al. 1991, J. Exp. Med. 173, 395-407; Sanz, 1991, J. Immunol. 147, 1720-1729.

Table 23. Germline V(D)J Segment Usage in Hybridoma Transcripts.

clone	h.c. CDR3	VH	DH	JH	l.c. CDR3	Vk	Jk
1E11	DITMVRGVPH (SEQ ID NO:209)	VH4- 34	DXP'1	JH4	QQYGSSPLT (SEQ ID NO:210)	VkA27/A11	Jk4
1G2	PANWNWYFVL (SEQ ID NO:211)	VH5- 51	DHQ52	JH2	QQFISYPQLT (SEQ ID NO:212)	VkL18	Jk4
6G5	VINWFDP (SEQ ID NO:213)	VH4- 34	n.d.	JH5	QQANSFPYT (SEQ ID NO:214)	VkL19	Jk2

10C5	VINWFDP (SEQ ID NO:213)	VH4- 34	n.d.	JH5	QQANSFPYT (SEQ ID NO:214)	VkL19	Jk2
4D1	DQLGLFDY (SEQ ID NO:215)	VH5- 51	DHQ52	JH4	QQYDSYPYT (SEQ ID NO:216)	VkL15	Jk2

n.d. could not be determined from nucleotide sequence.

Example 42.

Construction of Minigenes for Expression of Human IgGkappa AntiCD4 Antibodies in Transfected Cell lines.

This example demonstrates the process of making a wholly artificial gene that encodes an immunoglobulin polypeptide (i.e., an immunoglobulin heavy chain or light chain). Plasmids were constructed so that PCR amplified V heavy and V light chain cDNA sequences could be used to reconstruct complete heavy and light chain minigenes.

The kappa light chain plasmid, pCK7-96, includes the kappa constant region and polyadenylation site (SEQ ID NO:217), such that kappa sequences amplified with 5' primers that include HindIII sites upstream of the initiator methionine can be digested with HindIII and BbsI, and cloned into pCK7-96 digested with HindIII and BbsI to reconstruct a complete light chain coding sequence together with a polyadenylation site. This cassette can be isolated as a HindIII/NotI fragment and ligated to transcription promoter sequences to create a functional minigene for transfection into cells.

The gamma1 heavy chain plasmid, pCG7-96, includes the human gamma1 constant region and polyadenylation site (SEQ ID NO:218), such that gamma sequences amplified with 5' primers that include HindIII sites upstream of the initiator methionine can be digested with HindIII and AgeI, and cloned into pCG7-96 digested with HindIII and AgeI to reconstruct a complete gamma1 heavy chain coding sequence together with a

polyadenylation site. This cassette can be isolated as a HindIII/SalI fragment and ligated to transcription promoter sequences to create a functional minigene for transfection into cells.

The following example demonstrates how nucleotide sequence data from hybridomas can be used to reconstruct functional Ig heavy and light chain minigenes. The nucleotide sequences of heavy and light chain transcripts from hybridomas 6G5 and 10C5 were used to design an overlapping set of synthetic oligonucleotides to create synthetic V sequences with identical amino acid coding capacities as the natural sequences. The synthetic heavy and kappa light chain sequences (designated HC6G5 (SEQ ID NO:219) and LC6G5 (SEQ ID NO:220) differed from the natural sequences in three ways: strings of repeated nucleotide bases were interrupted to facilitate oligonucleotide synthesis and PCR amplification; optimal translation initiation sites were incorporated according to Kozak's rules (Kozak, 1991, J. Biol. Chem. 266, 19867-19870); and, HindIII sites were engineered upstream of the translation initiation sites.

A. Synthetic kappa light chain.

Light chain PCR reaction 1.

The following oligonucleotides were pooled: o-548 (SEQ ID NO:221), o-549 (SEQ ID NO:222), o-550 (SEQ ID NO:223), o-551 (SEQ ID NO:224), o-552 (SEQ ID NO:225), o-563 (SEQ ID NO:226), o-564 (SEQ ID NO:227), o-565 (SEQ ID NO:228), o-566 (SEQ ID NO:229), o-567 (SEQ ID NO:230), and amplified with the following 2 primers: o-527 (SEQ ID NO:231) and o-562 (SEQ ID NO:232).

Light chain PCR reaction 2.

The following oligonucleotides were pooled: o-553 (SEQ ID NO:233), o-554 (SEQ ID NO:234), o-555 (SEQ ID NO:235), o-556 (SEQ ID NO:236), o-557 (SEQ ID NO:237), o-558 (SEQ ID NO:238), o-559 (SEQ ID NO:239), o-560 (SEQ ID NO:240), o-561 (SEQ ID NO:241), o-562 (SEQ ID NO:232), and amplified with the

262

Light chain PCR reaction 3.

The product of light chain PCR reaction 3 was then digested with HindIII and BbsI and cloned into HindIII/BbsI digested pCK7-96 (SEQ ID NO:217) to generate pLC6G5 (SEQ ID NO:243).

Heavy chain PCR reaction 1.

Heavy chain PCR reaction 2.

Heavy chain PCR reaction 3.

The product of heavy chain reaction 3 was then digested with HindIII and AgeI and cloned into HindIII/AgeI digested pCG7-96 (SEQ ID NO:218) to generate pHCG5 (SEQ ID NO:268).

12030

Used in Minigene Construction

pCK7-96 (SEQ ID NO:217)

TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTTCGGCTGCGGCGAGCGGTATC
AGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA
TGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTC
CATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAA
CCCGACAGGACTATAAAGATAACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG
TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTT
TCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTTCGCTCCAAGCTGGGCTG
TGTGCACGAACCCCCGTTACGCCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGT
CCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGA
GCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAG
AAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTA
GCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAG
ATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGC
TCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCA
CCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACT
TGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCG
TTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCAT
CTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACC GGCTCCAGATTTATCAGCA
ATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCAT
CCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCCGCCAGTTAATAGTTTGCGCA
ACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTC
AGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAGCGGT
TAGCTCCTTCGGTCCTCCGATCGTTGTGAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGG
TTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACT
GGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCC
GGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAA
CCCACTCGTGACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGC
AAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATAC
TCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGA
TACATATTTGAATGTATTTAGAAAAATAACAAATAGGGGTTCGCGCACATTTCCCCGAAA
AGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTA
TCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAG
CTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGG
CGCGTCAGCGGGTGTGGCGGGTGTGCGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTG

TACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGC
ATCAGGCGCCATTTCGCCATTTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTC
TTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGC
CAGGGTTTTCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTAGCGGCCGCGGTC
CAACCACCAATCTCAAAGCTTGGTACCCGGGAGCCTGTTATCCCAGCACAGTCCTGGAAGAG
GCACAGGGGAAATAAAAGCGGACGGAGGCTTTCCTTGACTCAGCCGCTGCCTGGTCTTCTTC
AGACCTGTTCTGAATTCTAAACTCTGAGGGGGTTCGGATGACGTGGCCATTCTTTGCCTAAAG
CATTGAGTTTACTGCAAGGTCAGAAAAGCATGCAAAGCCCTCAGAATGGCTGCAAAGAGCTC
CAACAAAACAATTTAGAACTTTATTAAGGAATAGGGGGAAGCTAGGAAGAACTCAAAACAT
CAAGATTTTAAATACGCTTCTTGGTCTCCTTGCTATAATTATCTGGGATAAGCATGCTGTTT
TCTGTCTGTCCCTAACATGCCCTGTGATTATCCGCAAACAACACACCCAAGGGCAGAACTTT
GTTACTTAAACACCATCCTGTTTGCTTCTTTCCTCAGGAAGTGTGGCTGCACCATCTGTCTT
CATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAAGTGCCTCTGTTGTGTGCCTGCTGA
ATAACTTCTATCCCAGAGAGGCCAAAGTACAGTGGAAGGTGGATAACGCCCTCCAATCGGGT
AACTCCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACCTACAGCCTCAGCAGCAC
CCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCTACGCCTGCGAAGTCACCCATC
AGGGCCTGAGCTCGCCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTTAGAGGGAGAAGTGC
CCCCACCTGCTCCTCAGTTCAGCCTGACCCCTCCCATCCTTTGGCCTCTGACCCTTTTTTC
CACAGGGGACCTACCCCTATTGCGGTCTCCAGCTCATCTTTCACCTCACCCCCCTCCTCCT
CCTTGGCTTTAATTATGCTAATGTTGGAGGAGAATGAATAAATAAAGTGAATCTTTGCACCT
GTGGTTTCTCTCTTTCCTCAATTTAATAATTATTATCTGTTGTTTACCAACTACTCAATTTCT
TCTTATAAGGGACTAAATATGTAGTCATCCTAAGGCGCATAACCATTTATAAAAATCATCCT
TCATTCTATTTTACCCTATCATCCTCTGCAAGACAGTCCTCCCTCAAACCCACAAGCCTTCT
GTCCTCACAGTCCCCTGGGCCATGGATCCTCACATCCCAATCCGCGGCCGCAATTCGTAATC
ATGGTCATAGCTGTTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG
CCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCG
TTGCGCTCACTGCCCCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGG
CCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGC

pCG7-96 (SEQ ID NO:218)

GAACTCGAGCAGCTGAAGCTTTCTGGGGCAGGCCAGGCCTGACCTTGGCTTTGGGGCAGGGA
GGGGGCTAAGGTGAGGCAGGTGGCGCCAGCCAGGTGCACACCCAATGCCCATGAGCCCAGAC
ACTGGACGCTGAACCTCGCGGACAGTTAAGAACCCAGGGGCTCTGCGCCCTGGGCCCAGCT
CTGTCCCACACCGCGGTCACATGGCACCACCTCTCTTGCAGCCTCCACCAAGGGCCCATCGG
TCTTCCCCCTGGCACCCCTCCTCCAAGAGCACCTCTGGGGGCACAGCGGCCCTGGGCTGCCTG
GTCAAGGACTACTTCCCCGAACCGGTGACGGTGTCTGTGGAAGTCAAGGCGCCCTGACCAGCGG
CGTGCACACCTTCCCGGCTGTCTACAGTCCTCAGGACTCTACTCCCTCAGCAGCGTGGTGA
CCGTGCCCTCCAGCAGCTTGGGCACCCAGACCTACATCTGCAACGTGAATCACAAGCCCAGC

AACACCAAGGTGGACAAGAAAGTTGGTGAGAGGCCAGCACAGGGAGGGAGGGTGTCTGCTGG
AAGCCAGGCTCAGCGCTCCTGCCTGGACGCATCCCGGCTATGCAGCCCCAGTCCAGGGCAGC
AAGGCAGGCCCCGTCTGCCTCTTCACCCGGAGGCCTCTGCCCGCCCCACTCATGCTCAGGGA
GAGGGTCTTCTGGCTTTTTCCCCAGGCTCTGGGCAGGCACAGGCTAGGTGCCCTAACCCAG
GCCCTGCACACAAAGGGGCAGGTGCTGGGCTCAGACCTGCCAAGAGCCATATCCGGGAGGAC
CCTGCCCTGACCTAAGCCCACCCCAAAGGCCAAACTCTCCTCCTCAGCTCGGACACCT
TCTCTCCTCCCAGATTCCAGTAACTCCCAATCTTCTCTCTGCAGAGCCCAAATCTTGTGACA
AAACTCACACATGCCACCGTGCCAGGTAAGCCAGCCCAGGCCTCGCCCTCCAGCTCAAGG
CGGGACAGGTGCCCTAGAGTAGCCTGCATCCAGGGACAGGCCCCAGCCGGGTGCTGACACGT
CCACCTCCATCTCTTCCTCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCCTCTTCCCC
CCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGA
CGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATA
ATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTC
ACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGC
CCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGTGGGACCCGTGGGGTGCGAG
GGCCACATGGACAGAGGCCGGCTCGGCCCCACCCTCTGCCCTGAGAGTGACCGCTGTACCAAC
CTCTGTCCCTACAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGATG
AGCTGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATC
GCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACACTACAAGACCACGCCTCCCGTGCT
GGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGC
AGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGCACAACCACTACACGCAGAAG
AGCCTCTCCCTGTCTCCGGGTAAATGAGTGCGACGGCCGGCAAGCCCCGCTCCCCGGGCTC
TCGCGGTGCGACGAGGATGCTTGGCACGTACCCCTGTACATACTTCCCGGGCGCCCAGCAT
GGAAATAAAGCACCCAGCGCTGCCCTGGGCCCCTGCGAGACTGTGATGGTTCTTTCCACGGG
TCAGGCCGAGTCTGAGGCCTGAGTGGCATGAGGGAGGCAGAGCGGGTCCCCTGTCCCCACA
CTGGCCCAGGCTGTGCAGGTGTGCCTGGGCCCCCTAGGGTGGGGCTCAGCCAGGGGCTGCCC
TCGGCAGGGTGGGGGATTTGCCAGCGTGGCCCTCCCTCCAGCAGCACCTGCCCTGGGCTGGG
CCACGGGAAGCCCTAGGAGCCCCCTGGGGACAGACACAGCCCCCTGCCTCTGTAGGAGACTG
TCCTGTTCTGTGAGCGCCCCTGTCCTCCCGACCTCCATGCCCACTCGGGGGCATGCCTGCAG
GTGCACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCATCGATGATATCAGATCTGCCGG
TCTCCCTATAGTGAGTCGTATTAATTTGATAAGCCAGGTTAACCTGCATTAATGAATCGGC
CAACGCGCGGGGAGAGGCGGTTTGGGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTC
GCTGCGCTCGGTTCGTTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGT
TATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAGGCCAGCAAAGGCC
AGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCA
TCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
CGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATAC
CTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCT

CAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCG
ACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCG
CCTACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA
GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTC
TGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
GCTGGTAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCA
AGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTAAG
GGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGA
AGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAAT
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCG
TCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACC GGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGA
GCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAG
CTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATC
GTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTTCAGCTCCGGTTCCCAACGATCAAGGCG
AGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTG
TCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTT
ACTGTCATGCCATCCGTAAGATGCTTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTG
AGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC
CACATAGCAGAACTTTAAAAGTGCTCATCATTTGGAAAACGTTCTTCGGGGCGAAAACCTCTCA
AGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCCACTGATCTTC
AGCATCTTTTACTTTACACGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAA
AAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTAT
TGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAA
TAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCA
TTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTTCGTCTCGCGCGT
TTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCT
GTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTG
GGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGGACATA
TTGTTCGTTAGAACGCGGCTACAATTAATACATAACCTTATGTATCATAACATAACGATTTAG
GTGACACTATA

O-548 (SEQ ID NO:221)

ATGGTCCCAGCTCAGCTCCTCGGTCTCCTGCTGCTCTGGTTCCC

O-549 (SEQ ID NO:222)

AGGTTCCAGATGCGACATCCAGATGACCCAGTCTCCATCTTCCG

O-550 (SEQ ID NO:223)

TGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGTCGGGCG

O-551 (SEQ ID NO:224)

AGTCAGGATATTAGCAGCTGGTTAGCCTGGTATCAGCATAAACC

O-552 (SEQ ID NO:225)

AGGTAAAGCACCTAAGCTCCTGATCTATGCTGCATCCAGTTTGC

O-563 (SEQ ID NO:226)

AGGAGCTTAGGTGCTTTACCTGGTTTATGCTGATACCAGGCTAA

O-564 (SEQ ID NO:227)

CCAGCTGCTAATATCCTGACTCGCCCGACAAGTGATGGTGACTION

O-565 (SEQ ID NO:228)

TGTCTCCTACAGATGCAGACACGGAAGATGGAGACTGGGTCATC

O-566 (SEQ ID NO:229)

TGGATGTGCGATCTGGAACCTGGGAACCAGAGCAGCAGGAGACC

O-567 (SEQ ID NO:230)

GAGGAGCTGAGCTGGGACCATCATGGTGGCAAGCTTAGAGTC

O-527 (SEQ ID NO:231)

GACTCTAAGCTTGCCACCATGATGGTCC

O-562 (SEQ ID NO:232)

ACCTTGATGGGACACCACTTTGCAAACCTGGATGCAGCATAGATC

O-553 (SEQ ID NO:233)

AAAGTGGTGTCCCATCAAGGTTTCAGCGGAAGTGGATCTGGGACA

O-554 (SEQ ID NO:234)

GATTTCACTCTCACCATCAGCAGCCTGCAGCCTGAAGATTTTGC

O-555 (SEQ ID NO:235)

AACTTACTATTGTCAACAGGCTAATAGTTTCCCGTACACTTTTG

O-556 (SEQ ID NO:236)

GTCAGGGAACCAAGCTGGAGATCAAACGAACTGTGGCTGCACCA

O-557 (SEQ ID NO:237)

TCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGA

O-558 (SEQ ID NO:238)

GGGAAGATGAAGACAGATGGTGCAGCCACAGTTCGTTTGA

O-559 (SEQ ID NO:239)

TCTCCAGCTTGGTTCCTTGACCAAAAGTGTACGGGAAACTATTA

O-560 (SEQ ID NO:240)

GCCTGTTGACAATAGTAAGTTGCAAAATCTTCAGGCTGCAGGCT

O-561 (SEQ ID NO:241)

GCTGATGGTGAGAGTGAAATCTGTCCCAGATCCACTTCCGCTGA

O-493 (SEQ ID NO:242)

TCAACTGCTCATCAGATGGC

pLC6G5 (SEQ ID NO:243)

TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTTCGGCTGCGGCGAGCGGTATC
AGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA
TGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTC
CATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAA
CCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG
TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTT
TCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTTCGCTCCAAGCTGGGCTG
TGTGCACGAACCCCCCGTTACGCCCCGACCGCTGCGCCTTATCCGGTAACATCGTCTTGAGT
CCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGA
GCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAG
AAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTA
GCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAG
ATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGC
TCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTTCATGAGATTATCAAAAAGGATCTTCA
CCTAGATCCTTTTAAATTAAAAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAAACT

TGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCG
TTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCAT
CTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACC GGCTCCAGATTTATCAGCA
ATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCAT
CCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCCGCCAGTTAATAGTTTGCGCA
ACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTC
AGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAGCGGT
TAGCTCCTTCGGTCCTCCGATCGTTGTGAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGG
TTATGGCAGCACTGCATAATTCTCTTACTGTGTCATGCCATCCGTAAGATGCTTTTCTGTGACT
GGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCC
GGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAA
CCCACCTCGTGACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGC
AAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATAC
TCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGA
TACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAA
AGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAAATAGGCGTA
TCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAG
CTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGG
CGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACCTATGCGGCATCAGAGCAGATTG
TACTGAGAGTGACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGC
ATCAGGCGCCATTCGCCATTTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTC
TTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGC
CAGGGTTTTCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTAGCGGCCGCGGTC
CAACCACCAATCTCAAAGCTTGCCACCATGATGGTCCCAGCTCAGCTCCTCGGTCTCCTGCT
GCTCTGGTTCCCAGGTTCCAGATGCGACATCCAGATGACCCAGTCTCCATCTTCCGTGTCTG
CATCTGTAGGAGACAGAGTCACCATCACTTGTGCGGGCAGTCAGGATATTAGCAGCTGGTTA
GCCTGGTATCAGCATAAACCAGGTAAAGCACCTAAGCTCCTGATCTATGCTGCATCCAGTTT
GCAAAGTGGTGTCCCATCAAGGTTTCAGCGGAAGTGGATCTGGGACAGATTTCACTCTCACCA
TCAGCAGCCTGCAGCCTGAAGATTTTGCAACTTACTATTGTCAACAGGCTAATAGTTTCCCG
TACACTTTTGGTCAGGGAACCAAGCTGGAGATCAAACGAACTGTGGCTGCACCATCTGTCTT
CATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAAGTGCCTCTGTTGTGTGCCTGCTGA
ATAACTTCTATCCCAGAGAGGCCAAAGTACAGTGGAAGGTGGATAACGCCCTCCAATCGGGT
AACTCCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACCTACAGCCTCAGCAGCAC
CCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCTACGCCTGCGAAGTCACCCATC
AGGGCCTGAGCTCGCCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTTAGAGGGAGAAGTGC
CCCCACCTGCTCCTCAGTTCAGCCTGACCCCTCCCATCCTTTGGCCTCTGACCCTTTTTTC
CACAGGGGACCTACCCCTATTGCGGTCTCCAGCTCATCTTTCACCTACCCCCCTCCTCCT

CCTTGGCTTTAATTATGCTAATGTTGGAGGAGAATGAATAAATAAAGTGAATCTTTGCACCT
GTGGTTTCTCTCTTTCTCAATTTAATAATTATTATCTGTTGTTTACCAACTACTCAATTC
TCTTATAAGGGACTAAATATGTAGTCATCCTAAGGCGCATAACCATTTATAAAAATCATCCT
TCATTCTATTTTACCCTATCATCCTCTGCAAGACAGTCCTCCCTCAAACCCACAAGCCTTCT
GTCCTCACAGTCCCCTGGGCCATGGATCCTCACATCCCAATCCGCGGGCCGCAATTCGTAATC
ATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG
CCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCG
TTGCGCTCACTGCCCCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGG
CCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGC

O-528 (SEQ ID NO:244)

TTCTTCCTCCTCCTGGTGGCAGCTCCTAGATGGGTCCTGTCTC

O-529 (SEQ ID NO:245)

AGGTGCAGCTACAGCAGTGGGGCGCAGGACTGTTGAAGCCTTC

O-530 (SEQ ID NO:246)

GGAGACCCTGTCCCTCACCTGCGCTGTCTATGGTGGTTCCTTC

O-531 (SEQ ID NO:247)

AGTGGTTACTACTGGAGCTGGATCCGCCAGCCACCAGGTAAGG

O-532 (SEQ ID NO:248)

GTCTGGAGTGGATTGGTGAAATCAATCATAGTGGAAGCACCAA

O-543 (SEQ ID NO:249)

TTACCAATCCACTCCAGACCCTTACCTGGTGGCTGGCGGATC

O-544 (SEQ ID NO:250)

CAGCTCCAGTAGTAACCACTGAAGGAACCACCATAGACAGCGC

O-545 (SEQ ID NO:251)

AGGTGAGGGACAGGGTCTCCGAAGGCTTCAACAGTCCTGCGCC

O-546 (SEQ ID NO:252)

CCACTGCTGTAGCTGCACCTGAGACAGGACCCATCTAGGAGCT

O-547 (SEQ ID NO:253)

GCCACCAGGAGGAGGAAGAACCACAGGTGTTTCATGGTGGCAAGCTTG

O-496 (SEQ ID NO:254)
CATGAAACACCTGTGGTTCTTCC

O-542 (SEQ ID NO:255)
TCTTGAGAGACGGGTTGTAGTTGGTGCTTCCACTATGATTGAT

O-533 (SEQ ID NO:256)
CTACAACCCGTCTCTCAAGAGTCGAGTCACCATATCAGTAGAC

O-534 (SEQ ID NO:257)
ACGTCCAAGAACCAGTTCTCTCTGAAACTGAGCTCTGTGACCG

O-535 (SEQ ID NO:258)
CTGCGGACACGGCTGTGTATTACTGTGCGAGAGTAATTAATTG

O-536 (SEQ ID NO:259)
GTTTCGACCCTTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCA

O-537 (SEQ ID NO:260)
GCCTCAACCAAGGGCCCATCGGTCTTCCCCCTGGCACC

O-539 (SEQ ID NO:262)
CCTGGCCCCAAGGGTCGAACCAATTAATTACTCTCGCACAGTA

O-540 (SEQ ID NO:263)
ATACACAGCCGTGTCCGCAGCGGTCACAGAGCTCAGTTTCAGA

O-541 (SEQ ID NO:264)
GAGAACTGGTTCTTGGACGTGTCTACTGATATGGTGACTCGAC

O-538 (SEQ ID NO:261)
CGATGGGGCCCTTGGTTGAGGCTGAGGAGACGGTGACCAGGGTTC

O-490 (SEQ ID NO:265)
GAAGCACCAACTACAACCCG

O-520 (SEQ ID NO:266)
GAGTTCCACGACACCGTCACC

O-521 (SEQ ID NO:267)
GACCTCAAGCTTGCCACCATGAAACACCTGTGG

pHC6G5 (SEQ ID NO:268)
GAACTCGAGCAGCTGAAGCTTGCCACCATGAAACACCTGTGGTTCTTCCTCCTCCTGGTGGC
AGCTCCTAGATGGGTCCTGTCTCAGGTGCAGCTACAGCAGTGGGGCGCAGGACTGTTGAAGC
CTTCGGAGACCCTGTCCCTCACCTGCGCTGTCTATGGTGGTTCTTCAGTGGTTACTACTGG
AGCTGGATCCGCCAGCCACCAGGTAAGGGTCTGGAGTGGATTGGTGAAATCAATCATAGTGG
AAGCACCAACTACAACCCGTCTCTCAAGAGTCGAGTCACCATATCAGTAGACACGTCCAAGA
ACCAGTTCTCTCTGAACTGAGCTCTGTGACCGCTGCGGACACGGCTGTGTATTACTGTGCG
AGAGTAATTAATTGGTTCGACCCTTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGCCTC
AACCAAGGGCCCATCGGTCTTCCCCCTGGCACCTCCTCCAAGAGCACCTCTGGGGGCACAG
CGGCCCTGGGCTGCCTGGTCAAGGACTACTTCCCCGAACCGGTGACGGTGTCTGTGGAACCTCA
GGCGCCCTGACCAGCGGCTGCACACCTTCCCGGCTGTCCTACAGTCCTCAGGACTCTACTC
CCTCAGCAGCGTGGTGACCGTGCCCTCCAGCAGCTTGGGCACCCAGACCTACATCTGCAACG
TGAATCACAAGCCCAGCAACACCAAGGTGGACAAGAAAGTTGGTGAGAGGCCAGCACAGGGA
GGGAGGGTGTCTGCTGGAAGCCAGGCTCAGCGCTCCTGCCTGGACGCATCCCGGCTATGCAG
CCCCAGTCCAGGGCAGCAAGGCAGGCCCCGTCTGCCTCTTCACCCGGAGGCCTCTGCCCCGCC
CCACTCATGCTCAGGGAGAGGGTCTTCTGGCTTTTTCCCCAGGCTCTGGGCAGGCACAGGCT
AGGTGCCCTAACCCAGGCCCTGCACACAAAGGGGCAGGTGCTGGGCTCAGACCTGCCAAGA
GCCATATCCGGGAGGACCCTGCCCCTGACCTAAGCCCACCCCAAAGGCCAAACTCTCCTACTC
CCTCAGCTCGGACACCTTCTCTCCTCCCAGATTCCAGTAACTCCCAATCTTCTCTCTGCAGA
GCCCAAATCTTGTGACAAAACCTCACACATGCCACCGTGCCAGGTAAGCCAGCCCAGGCCT
CGCCCTCCAGCTCAAGGCGGGACAGGTGCCCTAGAGTAGCCTGCATCCAGGGACAGGCCCCA
GCCGGGTGCTGACACGTCCACCTCCATCTCTTCTCAGCACCTGAACTCCTGGGGGGACCGT
CAGTCTTCCTCTTCCCCCAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTC
ACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGA
CGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACC
GTGTGGTCAGCGTCTCACCGTCTGACACAGGACTGGCTGAATGGCAAGGAGTACAAGTGC
AAGGTCTCCAACAAAGCCCTCCCAGCCCCATCGAGAAAACCATCTCCAAGCCAAAGGTGG
GACCCGTGGGGTGCGAGGGCCACATGGACAGAGGCCGGCTCGGCCCACCCTCTGCCCTGAGA
GTGACCGCTGTACCAACCTCTGTCCCTACAGGGCAGCCCCGAGAACCACAGGTGTACACCCT
GCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTGACCTGCCTGGTCAAAGGCT
TCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAG
ACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCTCTACAGCAAGCTCACCGTGGA

CAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGCACA
ACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGCGACGGCCGGCAAGC
CCCCGCTCCCCGGGCTCTCGCGGTGCGACGAGGATGCTTGGCACGTACCCCCCTGTACATACT
TCCCGGGCGCCCAGCATGGAAATAAAGCACCCAGCGCTGCCCTGGGCCCCCTGCGAGACTGTG
ATGGTTCTTTCCACGGGTCAGGCCGAGTCTGAGGCCTGAGTGGCATGAGGGAGGCAGAGCGG
GTCCCACTGTCCCCACACTGGCCCAGGCTGTGCAGGTGTGCCTGGGCCCCCTAGGGTGGGGC
TCAGCCAGGGGCTGCCCTCGGCAGGGTGGGGGATTTGCCAGCGTGGCCCTCCCTCCAGCAGC
ACCTGCCCTGGGCTGGGCCACGGGAAGCCCTAGGAGCCCCCTGGGGACAGACACAGCCCCCT
GCCTCTGTAGGAGACTGTCCTGTTCTGTGAGCGCCCCCTGTCCTCCCGACCTCCATGCCCACT
CGGGGGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCATCGAT
GATATCAGATCTGCCGGTCTCCCTATAGTGAGTCGTATTAATTTTCGATAAGCCAGGTAAACC
TGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCT
TCCTCGCTCACTGACTCGCTGCGCTCGGTGCTTTCGGCTGCGGCGAGCGGTATCAGCTCACTC
AAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAA
AAGGCCAGCAAAAGGCCAGGAACCGTAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
CGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGG
ACTATAAAGATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCC
TGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGC
TCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGA
ACCCCCCGTTCAGCCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTAT
GTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGT
ATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGAT
CCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGC
AGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAA
CGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCC
TTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGAC
AGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCAT
AGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCA
GTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAG
CCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAA²TTTATCCGCCTCCATCCAGTCTAT
TAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTG
CCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTACGCTCCGGT
TCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTT
CGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAG
CACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTAC
TCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAAT
ACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTTGGAAAACGTTCTT

CGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGT
 GCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGG
 AAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCT
 TCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTT
 GAATGTATTTAGAAAAATAAACAAATAGGGGTTCGCGCACATTTCCCCGAAAAGTGCCACC
 TGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGC
 CCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAG
 ACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGC
 GGGTGTGGCGGGTGTGGGGCTGGCTTAACCTATGCGGCATCAGAGCAGATTGTACTGAGAG
 TGCACCATATGGACATATTGTCGTTAGAACGCGGCTACAATTAATACATAACCTTATGTATC
 ATACACATACGATTTAGGTGACACTATA

HC6G5 (SEQ ID NO:219)

AAGCTTGCCACCATGAAACACCTGTGGTTCTTCCTCCTCCTGGTGGCAGCTCCTAGATGGGT
 CCTGTCTCAGGTGCAGCTACAGCAGTGGGGCGCAGGACTGTTGAAGCCTTCGGAGACCCTGT
 CCCTCACCTGCGCTGTCTATGGTGGTTCTTCAGTGGTTACTACTGGAGCTGGATCCGCCAG
 CCACCAGGTAAGGGTCTGGAGTGGATTGGTGAAATCAATCATAGTGGAAGCACCAACTACAA
 CCCGTCTCTCAAGAGTCGAGTCACCATATCAGTAGACACGTCCAAGAACCAGTTCTCTCTGA
 AACTGAGCTCTGTGACCGCTGCGGACACGGCTGTGTATTACTGTGCGAGAGTAATTAATTGG
 TTCGACCCTTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGCCTCAACCAAGGGCCCATC
 GGTCTTCCCCCTGGCACCCCTCCTCCAAGAGCACCTCTGGGGGCACAGCGGCCCTGGGCTGCC
 TGGTCAAGGACTACTTCCCCGAACCGGT

LC6G5 (SEQ ID NO:220)

AAGCTTGCCACCATGATGGTCCCAGCTCAGCTCCTCGGTCTCCTGCTGCTCTGGTTCCCAGG
 TTCCAGATGCGACATCCAGATGACCCAGTCTCCATCTTCCGTGTCTGCATCTGTAGGAGACA
 GAGTCACCATCACTTGTGCGGGCGAGTCAGGATATTAGCAGCTGGTTAGCCTGGTATCAGCAT
 AAACCAGGTAAAGCACCTAAGCTCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGTGTCCC
 ATCAAGGTTTCAGCGGAAGTGGATCTGGGACAGATTTCACTCTCACCATCAGCAGCCTGCAGC
 CTGAAGATTTTGCAACTTACTATTGTCAACAGGCTAATAGTTTCCCGTACACTTTTGGTCAG
 GGAACCAAGCTGGAGATCAAACGAAGTGTGGCTGCACCATCTGTCTTC

Example 43.

Binding of Human Anti-CD4 Monoclonal Antibodies to Non-Human Primate Lymphocytes.

It is desirable to be able perform preclinical toxicology and pharmacokinetic studies of human anti-CD4 monoclonal antibodies in animal models. It is further

desirable for some purposes that the animal be a non-human primate that expresses CD4 comprising a cross-reactive epitope with human CD4 such that it is recognized by the monoclonal antibody. Three different non-human primate species, chimpanzee, rhesus, and cynomolgus monkeys, were tested for cross-reactive CD4 epitopes with the 5 different human anti-CD4 monoclonal antibodies from hybridomas 1E11, 1G2, 6G5, 10C5, and 4D1. Peripheral blood lymphocytes were isolated from whole blood of chimpanzee, rhesus, and cynomolgus monkeys. The isolated cells were double stained with human antibody from each of these 5 hybridomas (detected with FITC-anti-human IgG) and PE-anti-CD8 or PE-anti-CD4. The stained cells were then analyzed by flow cytometry to determine if each of the human monoclonal antibodies bound to endogenous CD4 on the surface of lymphocytes from each of these three non-human primates. Four of the five antibodies, 1E11, 6G5, 10C5, and 4D1, were found to bind to chimpanzee CD4 cells. Additionally, Four of the five antibodies, 6G5, 1G2, 10C5, and 4D1, were found to bind to both rhesus and cynomolgus CD4 cells. Thus, three of five antibodies, 6G5, 10C5, and 4D1, bind to CD4 cells in each of the three non-human primate species tested.

Example 44.

There are no known *in vitro* assays that can reliably predict whether a monoclonal antibody (mAb) will be nondepleting or immunosuppressive in patients. However, a correlation has been observed between the ability of three different mAbs to deplete (or not deplete) in humans and nonhuman primates such as chimpanzees and cynomolgus monkeys (See, e.g., M. Jonker *et al.*, *Clin. Exp. Immunol.*, 93:301-307 (1993); and J.A. Powelson *et al.*, *Transplantation*, 57:788-793 (1994)). Therefore a study was performed using human mAbs in nonhuman primates.

Chimpanzees were used in this study, because one of the anti-CD4 mAbs, 1E11, recognizes CD-4 only in chimpanzees and not in Rhesus or cynomolgus monkeys. A second mAb, 6G5, recognizes CD4 in chimpanzees, Rhesus and cynomolgus monkeys. A third mAb, 1G2, does not recognize CD4 in chimpanzees, but

does in Rhesus and cynomolgus monkeys. That mAb has already been shown to be nondepleting *in vivo* in cynomolgus monkeys.

In addition to examining the effect of human mAbs on CD4+ T cell numbers in peripheral blood, the effect of the mAb administration on *in vivo* T cell function was also evaluated. The most accepted manner to do this is to use animals that have been presensitized to an antigen such as tuberculin or tetanus toxoid and who will mount a hypersensitivity reaction in the skin.

Three male chimpanzees were enrolled in this study. Baseline whole blood samples were obtained on days -7, -3 and 1. After the blood draw on day 1, one chimpanzee each was intravenously infused with one of the two human mAbs (1E11 or 6G5) at 2 mg/kg. The third chimpanzee received an equal volume/kg of buffer only. Blood was drawn at 30 mins, 2 hrs, 8 hrs, 24 hrs and 48 hrs post-infusion. On day 2, a skin reactivity test was performed.

Results shown in Table 25 below clearly demonstrate that 1E11 caused transient depletion of peripheral lymphocytes, with most CD4+ T cells being depleted. Even though 6G5 did not cause lymphocyte or CD4+ T cell depletion, both mAbs were able to inhibit a hypersensitivity response to tetanus toxoid, compared to the control chimpanzee. Thus, both human mAbs appear to be immunosuppressive *in vivo*, and this immunosuppression does not necessarily require T cell depletion.

Table 25. Effect of Human mAbs on Peripheral Chimpanzee Lymphocytes

Peripheral Lymphocytes (million/ml)

Study Day	1E11	6G5	Control
-7	4.2	6.4	4.2
-4	4.0	9.9	4.4
1, pre-infusion	4.8	5.7	5.8
1, 30 min post	1.6	6.0	4.0
1, 2 hr post	1.0	6.7	5.2
1, 6 hr post	1.5	8.0	4.2
2	3.5	9.6	5.7
3	3.9	9.7	5.9

The foregoing description of the preferred embodiments of the present invention has been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise form disclosed, and many modifications and variations are possible in light of the above teaching. It will be apparent that certain changes and modifications may be practiced within the scope of the claims.

All publications and patent applications herein are incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference. Commonly assigned applications U.S.S.N. 08/544,404 filed 10 October 1995, U.S.S.N. 08/352,322, filed 7 December 1994, U.S.S.N. 08/209,741 filed 9 March 1994, U.S.S.N. 08/165,699 filed 10 December 1993 and U.S.S.N. 08/161,739 filed 03 December 1993, which is a continuation-in-part of 08/155,301 filed 18 November 1993, WO92/03918, USSN 07/810,279 filed 17 December 1991, USSN 07/853,408 filed 18 March 1992, USSN 07/904,068 filed 23 June 1992, USSN 07/990,860 filed 16 December 1992,

WO93/12227, and USSN 08/053,131 filed 26 April 1993 are each incorporated herein by reference.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Lonberg, Nils
Kay, Robert M.
- (ii) TITLE OF INVENTION: Transgenic Non-Human Animals for
Producing Heterologous Antibodies
- (iii) NUMBER OF SEQUENCES: 409
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Townsend and Townsend and Crew LLP
 - (B) STREET: Two Embarcadero Center, Eighth Floor
 - (C) CITY: San Francisco
 - (D) STATE: California
 - (E) COUNTRY: USA
 - (F) ZIP: 94111-3834
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/728,463
 - (B) FILING DATE: 10-OCT-1996
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/544,404
 - (B) FILING DATE: 10-OCT-1995
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/352,322
 - (B) FILING DATE: 07-DEC-1994
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/209,741
 - (B) FILING DATE: 09-MAR-1994
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/165,699
 - (B) FILING DATE: 10-DEC-1993
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/161,739
 - (B) FILING DATE: 03-DEC-1993
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/155,301
 - (B) FILING DATE: 18-NOV-1993
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/096,762
 - (B) FILING DATE: 22-JUL-1993

- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/053,131
 - (B) FILING DATE: 26-APR-1993
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/990,860
 - (B) FILING DATE: 16-DEC-1992
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/904,068
 - (B) FILING DATE: 23-JUN-1992
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/853,408
 - (B) FILING DATE: 18-MAR-1992
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/810,279
 - (B) FILING DATE: 17-DEC-1991
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/575,962
 - (B) FILING DATE: 31-AUG-1990
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/574,748
 - (B) FILING DATE: 29-AUG-1990
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: WO PCT/US91/06185
 - (B) FILING DATE: 29-AUG-1991
 - (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Serafini, Andrew T.
 - (B) REGISTRATION NUMBER: 41,303
 - (C) REFERENCE/DOCKET NUMBER: 014643-009020US
 - (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (415) 576-0200
 - (B) TELEFAX: (415) 576-0300
- (2) INFORMATION FOR SEQ ID NO:1:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTAADTGGGG

10

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Asp Ala Phe Asp Ile
1 5

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Asp Tyr Phe Asp Tyr
1 5

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Gly Ala Phe Asp Ile
1 5

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Lys Glu Arg Val
1

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Asn Asp Ser Val
1

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AAAGAAAGAG UU

12

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AACGACAGCG UU

12

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GAGCTGAGCT GGGGT

15

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GAGCTGAGCT GAGCTGGGGT

20

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAGCTGAGCT GAGCTGAGCT GGGGT

25

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAGCTGAGCT GAGCTGAGCT GAGCTGGGGT

30

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GGGGT

35

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 40 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGGGGT

40

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 45 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GGGGT

45

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:
GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGGGGT 50

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 55 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:
GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GGGGT 55

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:
GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGGGGT 60

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:
GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT 60
GGGGT 65

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 70 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT 60
GAGCTGGGGT 70

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 75 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT 60
GAGCTGAGCT GGGGT 75

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 80 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT 60
GAGCTGAGCT GAGCTGGGGT 80

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 85 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT 60
GAGCTGAGCT GAGCTGAGCT GGGGT 85

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 90 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT GAGCTGAGCT 60
GAGCTGAGCT GAGCTGAGCT GAGCTGGGGT 90

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

AATTGCGGCC GC 12

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CTTGAGCCCG CCTAATGAGC GGGCTTTTTT TTGCATACTG CGGCC 45

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCAATGGCCT GGATCCATGG CGCGCTAGCA TCGATATCTA GAGCTCGAGC A

51

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TGCAGATCTG AATTCCCGGG TACCAAGCTT ACGCGTACTA GTGCGGCCGC T

51

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 45 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

AATTAGCGGC CGCACTAGTA CGCGTAAGCT TGGTACCCGG GAATT

45

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CAGATCTGCA TGCTCGAGCT CTAGATATCG ATGCTAGCGC GCCATGGATC C

51

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

AGGCCATTGC GGCCGCAGTA TGCAAAAAAA AGCCCGCTCA TTAGGCGGGC T

51

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CGCGTGGCCG CAATGGCCA

19

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CTAGTGGCCA TTGCGGCCA

19

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CAGGATCCAG ATATCAGTAC CTGAAACAGG GCTTGC

36

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 36 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GAGCATGCAC AGGACCTGGA GCACACACAG CCTTCC

36

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 42 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GGACTGTGTC CCTGTGTGAT GCTTTTGATG TCTGGGGCCA AG

42

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 42 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CACCAAGTTG ACCTGCCTGG TCACAGACCT GACCACCTAT GA

42

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CCTGTGGACC ACCGCCTCCA CCTTCATCGT CCTCTTCCTC CT

42

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TGAGCCACGA AGACCCTGAG GTCAAGTTCA ACTGGTACGT GG

42

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TGGTATTACT ATGGTTCGGG GAGTTATTAT AACCACAGTG TC

42

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCCTGAAATG GAGCCTCAGG GCACAGTGGG CACGGACACT GT

42

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GCAGGGAGGA CATGTTTAGG ATCTGAGGCC GCACCTGACA CC

42

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GATCCTGGTT TAGTTAAAGA GGATTTTATT CACCCCTGTG TC

42

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GATCCAAGCA GT

12

(2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:
CTAGACTGCT TG 12

(2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:
CGCGTCGAAC TA 12

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:
AGCTTAGTTC GA 12

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 42 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:
GAATGGGAGT GAGGCTCTCT CATAACCTAT TCAGAACTGA CT 42

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 42 base pairs
 (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GAACTGTGGC TGCACCATCT GTCTTCATCT TCCCGCCATC TG

42

(2) INFORMATION FOR SEQ ID NO:50:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GAGGTACACT GACATACTGG CATGCCCCCC CCCCCC

36

(2) INFORMATION FOR SEQ ID NO:51:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 60 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GTACGCCATA TCAGCTGGAT GAAGTCATCA GATGGCGGGA AGATGAAGAC AGATGGTGCA

60

(2) INFORMATION FOR SEQ ID NO:52:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

TCATCAGATG GCGGGAAGAT GAAGACAGAT GGTGCA

36

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GTACGCCATA TCAGCTGGAT GAAG

24

(2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GAGGTACACT GACATACTGG CATG

24

(2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 60 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

GTACGCCATA TCAGCTGGAT GAAGACAGGA GACGAGGGGG AAAAGGGTTG GGGCGGATGC

60

(2) INFORMATION FOR SEQ ID NO:56:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

296

ACAGGAGACG AGGGGGAAAA GGGTTGGGGC GGATGC

36

(2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GTACTCCATA TCAGCTGGAT GAAG

24

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GGCTGATGCT GCACCAACTG TATCCATCTT CCCACCATCC AG

42

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CTCACGTTTCG GTGCTGGGAC CAAGCTGGAG CTGAAACGTA AG

42

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

ACTATGCTAT GGACTACTGG GGTCAAGGAA CCTCAGTCAC CG

42

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GGCCGCTCGA CGATAGCCTC GAGGCTATAA ATCTAGAAGA ATTCCAGCAA AGCTTTGGC

59

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CAAGAGCCCG CCTAATGAGC GGGCTTTTTT TTGCATACTG CGGCCGCT

48

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

AATTAGCGGC CGCAGTATGC AAAAAAAGC CCGCTCATTG GCGGGGCT

48

(2) INFORMATION FOR SEQ ID NO:64:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 59 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GCGGCCGCCT CGAGATCACT ATCGATTAAT TAAGGATCCA GCAGTAAGCT TCGGCCCGC 59

(2) INFORMATION FOR SEQ ID NO:65:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 53 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

GCGGCCGCAT CCCGGGTCTC GAGGTCGACA AGCTTTCGAG GATCCGCGGC CGC 53

(2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 47 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GCGGCCGCTG TCGACAAGCT TATCGATGGA TCCTCGAGTG CGGCCGC 47

(2) INFORMATION FOR SEQ ID NO:67:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 65 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GCGGCCGCTG TCGACAAGCT TCGAATTCAG ATCGATGTGG TACCTGGATC CTCGAGTGCG 60

GCCGC

65

(2) INFORMATION FOR SEQ ID NO:68:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GGCCGCAAGC TTACTGCTGG ATCCTTAATT AATCGATAGT GATCTCGAGG C

51

(2) INFORMATION FOR SEQ ID NO:69:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

GGCCGCCTCG AGATCACTAT CGATTAATTA AGGATCCAGC AGTAAGCTTG C

51

(2) INFORMATION FOR SEQ ID NO:70:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

CTCCAGGATC CAGATATCAG TACCTGAAAC AGGGCTTG C

39

(2) INFORMATION FOR SEQ ID NO:71:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

300

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

CTCGAGCATG CACAGGACCT GGAGCACACA CAGCCTTCC

39

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3618 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..3618
- (D) OTHER INFORMATION: /note= "vector pGPe"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

AATTAGCGGC CGCCTCGAGA TCACTATCGA TTAATTAAGG ATCCAGATAT CAGTACCTGA	60
AACAGGGCTG CTCACAACAT CTCTCTCTCT GTCTCTCTGT CTCTGTGTGT GTGTCTCTCT	120
CTGTCTCTGT CTCTCTCTGT CTCTCTGTCT CTGTGTGTGT CTCTCTCTGT CTCTCTCTCT	180
GTCTCTCTGT CTCTCTGTCT GTCTCTGTCT CTGTCTCTGT CTCTCTCTCT CTCTCTCTCT	240
CTCTCTCTCT CTCTCTCACA CACACACACA CACACACACA CACACCTGCC GAGTGACTCA	300
CTCTGTGCAG GGTGCGCCCT CGGGGCACAT GCAAATGGAT GTTTGTTCCA TGCAGAAAAA	360
CATGTTTCTC ATTCTCTGAG CCAAAAATAG CATCAATGAT TCCCCACCC TGCAGCTGCA	420
GGTTCACCCC ACCTGGCCAG GTTGACCAGC TTTGGGGATG GGGCTGGGGG TTCCATGACC	480
CCTAACGGTG ACATTGAATT CAGTGTTTTT CCATTTATCG AACTGCTGG AATCTGACCC	540
TAGGAGGGAA TGACAGGAGA TAGGCAAGGT CCAAACACCC CAGGGAAGTG GGAGAGACAG	600
GAAGGCTGTG TGTGCTCCAG GTCCTGTGCA TGCTGCAGAT CTGAATTCCC GGGTACCAAG	660
CTTGCGGCCG CAGTATGCAA AAAAAAGCCC GCTCATTAGG CGGGCTCTTG GCAGAACATA	720
TCCATCGCGT CCGCCATCTC CAGCAGCCGC ACGCGGCGCA TCTCGGGCAG CGTTGGGTCC	780
TGGCCACGGG TGCGCATGAT CGTGCTCCTG TCGTTGAGGA CCCGGCTAGG CTGGCGGGGT	840
TGCCTTACTG GTTAGCAGAA TGAATCACCG ATACGCGAGC GAACGTGAAG CGACTGCTGC	900
TGCAAAACGT CTGCGACCTG AGCAACAACA TGAATGGTCT TCGGTTTCCG TGTTTCGTAA	960

AGTCTGAAAA	CGCGGAAGTC	AGCGCCCTGC	ACCATTATGT	TCCGGATCTG	CATCGCAGGA	1020
TGCTGCTGGC	TACCCTGTGG	AACACCTACA	TCTGTATTAA	CGAAGCGCTG	GCATTGACCC	1080
TGAGTGATTT	TTCTCTGGTC	CCGCCGCATC	CATACCGCCA	GTTGTTTACC	CTCACAACGT	1140
TCCAGTAACC	GGGCATGTTT	ATCATCAGTA	ACCCGTATCG	TCACGATCCT	CTCTCGTTTC	1200
ATCGGTATCA	TTACCCCCAT	GAACAGAAAT	TCCCCCTTAC	ACGGAGGCAT	CAAGTGACCA	1260
AACAGGAAAA	AACCGCCCTT	AACATGGCCC	GCTTTATCAG	AAGCCAGACA	TTAACGCTTC	1320
TGGAGAAACT	CAACGAGCTG	GACGCGGATG	AACAGGCAGA	CATCTGTGAA	TCGCTTCACG	1380
ACCACGCTGA	TGAGCTTTAC	CGCAGCTGCC	TCGCGCGTTT	CGGTGATGAC	GGTGAAAACC	1440
TCTGACACAT	GCAGCTCCCG	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	1500
GACAAGCCCC	TCAGGGCGCG	TCAGCGGGTG	TTGGCGGGTG	TCGGGGCGCA	GCCATGACCC	1560
AGTCACGTAG	CGATAGCGGA	GTGTATACTG	GCTTAACTAT	GCGGCATCAG	AGCAGATTGT	1620
ACTGAGAGTG	CACCATATGC	GGTGTGAAAT	ACCGCACAGA	TGCGTAAGGA	GAAAATACCG	1680
CATCAGGCGC	TCTTCCGCTT	CCTCGCTCAC	TGACTCGCTG	CGCTCGGTCT	TTGGGCTGCG	1740
GCGAGCGGTA	TCAGCTCACT	CAAAGGCGGT	AATACGGTTA	TCCACAGAAT	CAGGGGATAA	1800
CGCAGGAAAG	AACATGTGAG	CAAAAGGCCA	GCAAAGGCC	AGGAACCGTA	AAAAGGCCGC	1860
GTTGCTGGCG	TTTTTCCATA	GGCTCCGCCC	CCCTGACGAG	CATCACAAAA	ATCGACGCTC	1920
AAGTCAGAGG	TGGCGAAACC	CGACAGGACT	ATAAAGATAC	CAGGCGTTTC	CCCCTGGAAG	1980
CTCCCTCGTG	CGCTCTCCTT	AGGTATCTCA	GTTCCGGTGT	GGTCGTTTCG	TCCAAGCTGG	2040
GCTGTGTGCA	CGAACCCCCC	GTTTCAGCCG	ACCGCTGCGC	CTTATCCGGT	AACTATCGTC	2100
TTGAGTCCAA	CCCGGTAAGA	CACGACTTAT	CGCCACTGGC	AGCAGCCACT	GGTAACAGGA	2160
TTAGCAGAGC	GAGGTATGTA	GGCGGTGCTA	CAGAGTTCTT	GAAGTGGTGG	CCTAACTACG	2220
GCTACACTAG	AAGGACAGTA	TTTGGTATCT	GCGCTCTGCT	GAAGCCAGTT	ACCTTCGGAA	2280
AAAGAGTTGG	TAGCTCTTGA	TCCGGCAAAC	AAACCACCGC	TGGTAGCGGT	GGTTTTTTTTG	2340
TTTGCAAGCA	GCAGATTACG	CGCAGAAAAA	AAGGATCTCA	AGAAGATCCT	TTGATCTTTT	2400
CTACGGGGTC	TGACGCTCAG	TGGAACGAAA	ACTCACGTTA	AGGGATTTTG	GTCATGAGAT	2460
TATCAAAAAG	GATCTTCACC	TAGATCCTTT	TAAATTAAAA	ATGAAGTTTT	AAATCAATCT	2520
AAAGTATATA	TGAGTAAACT	TGGTCTGACA	GTTACCAATG	CTTAATCAGT	GAGGCAGGTA	2580
TCTCAGCGAT	CTGTCTATTT	CGTTCATCCA	TAGTTGCCCT	ACTCCCCGTC	GTGTAGATAA	2640
CTACGATACG	GGAGGGCTTA	CCATCTGGCC	CCAGTGCTGC	AATGATACCG	CGAGACCCAC	2700
GCTCACCGGC	TCCAGATTTA	TCAGCAATAA	ACCAGCCAGC	CGGAAGGGCC	GAGCGCAGAA	2760
GTGGTCCTGC	AACTTTATCC	GCCTCCATCC	AGTCTATTAA	TTGTTGCCGG	GAAGCTAGAG	2820

TAAGTAGTTC GCCAGTTAAT AGTTTGCGCA ACGTTGTTGC CATTGCTGCA GGCATCGTGG	2880
TGTCACGCTC GTCGTTTGGT ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG	2940
TTACATGATC CCCCATGTTG TGCAAAAAAAG CGGTTAGCTC CTTCGGTCCT CCGATCGTTG	3000
TCAGAAGTAA GTTGGCCGCA GTGTTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC	3060
TTACTGTCAT GCCATCCGTA AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT	3120
TCTGAGAATA GTGTATGCGG CGACCGAGTT GCTCTTGCCC GGCCTCAACA CGGGATAATA	3180
CCGCGCCACA TAGCAGAACT TAAAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA	3240
AACTCTCAAG GATCTTACCG CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA	3300
ACTGATCTTC AGCATCTTTT ACTTTCACCA GCGTTTCTGG GTGAGCAAAA ACAGGAAGGC	3360
AAAAATGCCG AAAAAAGGGA ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC	3420
TTTTTCAATA TTATTGAAGC ATTTATCAGG GTTATTGTCT GATGAGCGGA TACATATTTG	3480
AATGTATTTA GAAAAATAAA CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC	3540
CTGACGTCTA AGAAACCATT ATTATCATGA CATTAACCTA TAAAAATAGG CGTATCACGA	3600
GGCCCTTTTCG TCTTCAAG	3618

(2) INFORMATION FOR SEQ ID NO:73:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

GGACTGTGTC CCTGTGTGAT GCTTTTGATG TCTGGGGCCA AG	42
--	----

(2) INFORMATION FOR SEQ ID NO:74:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

CACCAAGTTG ACCTGCCTGG TCACAGACCT GACCACCTAT GA	42
--	----

(2) INFORMATION FOR SEQ ID NO:75:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

TGGTATTACT ATGGTTCGGG GAGTTATTAT AACCACAGTG TC

42

(2) INFORMATION FOR SEQ ID NO:76:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

CAGCAGGTGC ACACCCAATG CCCATGAGCC CAGACACTGG AC

42

(2) INFORMATION FOR SEQ ID NO:77:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

TGAGCCCAGA CACTGGAC

18

(2) INFORMATION FOR SEQ ID NO:78:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 45 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

GTTAAAGAGG ATTTTATTCA CCCCTGTGTC CTCTCCACAG GTGTC

45

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 812 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(241..286, 373..677)
- (D) OTHER INFORMATION: /note= "human V-HI family gene V-H49.8"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

TTCCTCAGGC AGGATTTAGG GCTTGGTCTC TCAGCATCCC ACACTTGTAC AGCTGATGTG	60
GCATCTGTGT TTTCTTTCTC ATCCTAGATC AAGCTTTGAG CTGTGAAATA CCCTGCCTCA	120
TGAATATGCA AATAATCTGA GGTCTTCTGA GATAAATATA GATATATTGG TGCCCTGAGA	180
GCATCACATA ACAACCAGAT TCCTCCTCTA AAGAAGCCCC TGGGAGCACA GCTCATCACC	240
ATG GAC TGG ACC TGG AGG TTC CTC TTT GTG GTG GCA GCA GCT ACA G	286
Met Asp Trp Thr Trp Arg Phe Leu Phe Val Val Ala Ala Ala Thr	
1 5 10 15	
GTAAGGGGCT TCCTAGTCCT AAGGCTGAGG AAGGGATCCT GGTTTAGTTA AAGAGGATTT	346
TATTCACCCC TGTGTCCTCT CCACAG GT GTC CAG TCC CAG GTC CAG CTG GTG	398
Gly Val Gln Ser Gln Val Gln Leu Val	
20	
CAG TCT GGG GCT GAG GTG AAG AAG CCT GGG TCC TCG GTG AAG GTC TCC	446
Gln Ser Gly Ala Glu Val Lys Lys Pro Gly Ser Ser Val Lys Val Ser	
25 30 35 40	
TGC AAG GCT TCT GGA GGC ACC TTC AGC AGC TAT GCT ATC AGC TGG GTG	494
Cys Lys Ala Ser Gly Gly Thr Phe Ser Ser Tyr Ala Ile Ser Trp Val	
45 50 55	
CGA CAG GCC CCT GGA CAA GGG CTT GAG TGG ATG GGA AGG ATC ATC CCT	542
Arg Gln Ala Pro Gly Gln Gly Leu Glu Trp Met Gly Arg Ile Ile Pro	
60 65 70	
ATC CTT GGT ATA GCA AAC TAC GCA CAG AAG TTC CAG GGC AGA GTC ACG	590
Ile Leu Gly Ile Ala Asn Tyr Ala Gln Lys Phe Gln Gly Arg Val Thr	
75 80 85	
ATT ACC GCG GAC AAA TCC ACG AGC ACA GCC TAC ATG GAG CTG AGC AGC	638
Ile Thr Ala Asp Lys Ser Thr Ser Thr Ala Tyr Met Glu Leu Ser Ser	

90	95	100	
CTG AGA TCT GAG GAC ACG GCC GTG TAT TAC TGT GCG AGA GACACAGTGT			687
Leu Arg Ser Glu Asp Thr Ala Val Tyr Tyr Cys Ala Arg			
105	110	115	
GAAAACCCAC ATCCTGAGAG TGTCAGAAAC CCTGAGGGAG AAGGCAGCTG TGCCGGGCTG			747
AGGAGATGAC AGGGTTTATT AGGTTTAAGG CTGTTTACAA AATGGGTTAT ATATTTGAGA			807
AAAAA			812

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 117 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

Met	Asp	Trp	Thr	Trp	Arg	Phe	Leu	Phe	Val	Val	Ala	Ala	Ala	Thr	Gly
1				5					10					15	
Val	Gln	Ser	Gln	Val	Gln	Leu	Val	Gln	Ser	Gly	Ala	Glu	Val	Lys	Lys
			20					25					30		
Pro	Gly	Ser	Ser	Val	Lys	Val	Ser	Cys	Lys	Ala	Ser	Gly	Gly	Thr	Phe
			35				40					45			
Ser	Ser	Tyr	Ala	Ile	Ser	Trp	Val	Arg	Gln	Ala	Pro	Gly	Gln	Gly	Leu
		50				55					60				
Glu	Trp	Met	Gly	Arg	Ile	Ile	Pro	Ile	Leu	Gly	Ile	Ala	Asn	Tyr	Ala
65					70					75					80
Gln	Lys	Phe	Gln	Gly	Arg	Val	Thr	Ile	Thr	Ala	Asp	Lys	Ser	Thr	Ser
				85					90					95	
Thr	Ala	Tyr	Met	Glu	Leu	Ser	Ser	Leu	Arg	Ser	Glu	Asp	Thr	Ala	Val
			100					105					110		
Tyr	Tyr	Cys	Ala	Arg											
		115													

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

306

CCGGTCGACC GG

12

(2) INFORMATION FOR SEQ ID NO:82:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 40 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

CTAGCTCGAG TCCAAGGAGT CTGTGCCGAG GTGCAGCTGN

40

(2) INFORMATION FOR SEQ ID NO:83:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 40 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

GTTGCTCGAG TGAAAGGTGT CCAGTGTGAG GTGCAGCTGN

40

(2) INFORMATION FOR SEQ ID NO:84:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

GGCGCTCGAG TTCCACGACA CCGTCACCGG TTC

33

(2) INFORMATION FOR SEQ ID NO:85:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

CCTGCTCGAG GCAGCCAACG GCCACGCTGC TCG

33

(2) INFORMATION FOR SEQ ID NO:86:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 88 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

TACTGTGCGA GACGGCTAAA CTGGGGTTGA TGCTTTTGAT ATCTGGGGCC AAGGGACAAT

60

GGTCACCGTC TCTTCAGCCT CCACCAAG

88

(2) INFORMATION FOR SEQ ID NO:87:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 87 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

TACTGTGCGA GACACCGTAT AGCAGCAGCT GGCTTTGACT ACTGGGGCCA GGGAACCCTG

60

GTCACCGTCT CCTCAGCCTC CACCAAG

87

(2) INFORMATION FOR SEQ ID NO:88:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 87 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

TACTGTGCGA GATATTACTA CTACTACTAC GGTATGGACG TCTGGGGCCA AGGGACCACG

60

GTCACCGTCT CCTCAGCCTC CACCAAG

87

(2) INFORMATION FOR SEQ ID NO:89:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 104 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

TACTGTGCGA GACATTACGA TATTTTGACT GGCCTACTA CTACTACGGT ATGGACGTCT	60
GGGGCCAAGG GACCACGGTC ACCGTCTCCT CAGCCTCCAC CAAG	104

(2) INFORMATION FOR SEQ ID NO:90:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 101 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

TACTGTGCGA GACGGAGGTA CTATGGTTCG GGGAGTTATT ATAACGTCTT TGACTACTGG	60
GGCCAGGGAA CCTGGTCACC GTCTCCTCAG CCTCCACCAA G	101

(2) INFORMATION FOR SEQ ID NO:91:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 84 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

TACTGTGCGA GACGGGGGGT GTCTGATGCT TTTGATATCT GGGGCAAGG GACAATGGTC	60
ACCGTCTCTT CAGCCTCCAC CAAG	84

(2) INFORMATION FOR SEQ ID NO:92:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 78 base pairs
 - (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

TACTGTGCGA GAGCAACTGG CGCTTTTGAT ATCTGGGGCC AAGGGACAAT GGTCACCGTC 60
TCTTCAGGGA GTGCATCC 78

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 99 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

TACTGTGCGA GATCGGCTAA CTGGGGATCC TACTACTACT ACGGTATGGA CGTCTGGGGC 60
CAAGGGACCA CGGTCACCGT CTCCTCAGGG AGTGCATCC 99

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

TACTGTGCGA GATACTTCCA GCACTGGGGC CAGGGCACCC TGGTCACCGT CTCCTCAGGG 60
AGTGCGTCC 69

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 81 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

TACTGTGCGA GACACGTAGC TAACTCTTTT GACTACTGGG GCCAGGGAAC CCTGGTCACC 60
 GTCTCCTCAG GGAGTGCATC C 81

(2) INFORMATION FOR SEQ ID NO:96:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 93 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

TACTGTGCGA GACAAATTAC TATGGTTCGG GGAGTTCCTT TTGACTACTG GGGCCAGGGA 60
 ACCCTGGTCA CCGTCTCCTC AGGGAGTGCA TCC 93

(2) INFORMATION FOR SEQ ID NO:97:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 72 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

TACTGTGCGA GACAATACTT CCAGCACTGG GGCCAGGGCA CCCTGGTCAC CGTCTCCTCA 60
 GGGAGTGCAT CC 72

(2) INFORMATION FOR SEQ ID NO:98:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 93 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

TACTGTGCGA GACAAACTGG GGACTACTAC TACTACGGTA TGGACGTCTG GGGCCAAGGG 60
 ACCACGGTCA CCGTCTCCTC AGGGAGTGCA TCC 93

(2) INFORMATION FOR SEQ ID NO:99:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 107 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

TACTGTGCGA GACATTACTA TGGTTCGGGG AGTTATGACT ACTACTACTA CGGTATGGAC 60
GTCTGGGGCC AAGGGACCAG GTCACCGTCT CCTCAGGGAG TGCATCC 107

(2) INFORMATION FOR SEQ ID NO:100:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 64 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

TACTGTGCGA GACAGGGAGT GGGGCCAGGG AACCCCTGGTC ACCGTCTCCT CAGCCTCCAC 60
CAAG 64

(2) INFORMATION FOR SEQ ID NO:101:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 80 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

TACTGTGCGA GATTCTGGGA GACTGGTTCG ACCCCTGGGG CCAGGGAACC CTGGTCACCG 60
TCTCCTCAGG GAGTGCATCC 80

(2) INFORMATION FOR SEQ ID NO:102:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 102 base pairs
 - (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

TACTGTGCGA GACGGAGGTA CTATGGTTCG GGGAGTTATT ATAACGTCTT TGACTACTGG 60
 GGCCAGGGAA CCCTGGTCAC CGTCTCCTCA GCCTCCACCA AG 102

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 78 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

TACTGTGCGA GACAAACCTG GGGAGGAGAC TACTGGGGCC AGGGAACCCT GGTCACCGTC 60
 TCCTCAGCCT CCACCAAG 78

(2) INFORMATION FOR SEQ ID NO:104:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 99 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

TACTGTGCGA GAGGATATAG TGGCTACGAT AACTACTACT ACGGTATGGA CGTCTGGGGC 60
 CAAGGGACCA CGGTCACCGT CTCCTCAGCC TCCACCAAG 99

(2) INFORMATION FOR SEQ ID NO:105:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 84 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

TACTGTGCGA GACAAACTGG GGAGGACTAC TTTGACTACT GGGGCCAGGG AACCTGGTC 60
ACCGTCTCCT CAGGGAGTGC ATCC 84

(2) INFORMATION FOR SEQ ID NO:106:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 94 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

TACTGTGCGA GATATAGTGG CTACGATTAC CTACTGGTAC TTCGATCTCT GGGGCCGTGG 60
CACCTGGTC ACCGTCTCCT CAGCCTCCAC CAAG 94

(2) INFORMATION FOR SEQ ID NO:107:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

TACTGTGCGA GAGCATCCCT CCCCTCCTTT GACTACTACG GTATGGACGT CTGGGGCCAA 60
GGGACCACGG TCACCGTCTC CTCAGCTCC ACCAAG 96

(2) INFORMATION FOR SEQ ID NO:108:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 76 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

TACTGTGCGA GACGGGGTGG GGTITGACTA CTGGGGCCAG GGAACCCTGG TCACCGTCTC 60
CTCAGGGAGT GCATCC 76

(2) INFORMATION FOR SEQ ID NO:109:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 94 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

TACTGTGTGC CGGTCGAAAC TTTACTACTA CTACTACGGT ATGGACGTCT GGGGCCAAGG 60
GACCACGGTC ACCGTCTCCT CAGGGAGTGC ATCC 94

(2) INFORMATION FOR SEQ ID NO:110:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 84 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

TACTGTGCGA GAGATATTTT GACTGGTTAA CGTGACTACT GGGGCCAGGG AACCTGGTC 60
ACCGTCTCCT CAGGGAGTGC ATCC 84

(2) INFORMATION FOR SEQ ID NO:111:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 93 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

TACTGTGCGA GACATGGTAT AGCAGCAGCT GGTACTGCTT TTGATATCTG GGGCCAAGGG 60
ACAATGGTCA CCGTCTCTTC AGGGAGTGCA TCC 93

(2) INFORMATION FOR SEQ ID NO:112:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 84 base pairs
 - (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

TACTGTGTGA GATCAACTGG GGTGATGCT TTTGATATCT GGGGCCAAGG GACAATGGTC 60
ACCGTCTCTT CAGGGAGTGC ATCC 84

(2) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 85 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

TACTGTGCGG AAATAGCAGC AGCTGCCCTA CTTGACTAC TGGGGCCAGG GAACCCTGGT 60
CACCGTCTCC TCAGGGAGTG CATCC 85

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 87 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

TACTGTGTGT GTATAGCAGC AGCTGGTAAA GGAAACGGCT ACTGGGGCCA GGGAACCCTG 60
GTCACCGTCT CCTCAGGGAG TGCATCC 87

(2) INFORMATION FOR SEQ ID NO:115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 75 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

TACTGTGCGA GACAAACTG GGGTGACTAC TGGGGCCAGG GAACCCTGGT CACCGTCTCC 60
 TCAGGGAGTG CATCC 75

(2) INFORMATION FOR SEQ ID NO:116:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

Tyr Cys Ala Arg Arg Leu Thr Gly Val Asp Ala Phe Asp Ile Trp Gly
 1 5 10 15
 Gln Gly Thr Met Val Thr Met Ser Ser Ala Ser Thr Lys
 20 25

(2) INFORMATION FOR SEQ ID NO:117:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

Tyr Cys Ala Arg His Arg Ile Ala Ala Ala Gly Phe Asp Tyr Trp Gly
 1 5 10 15
 Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys
 20 25

(2) INFORMATION FOR SEQ ID NO:118:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Tyr Cys Ala Arg Tyr Tyr Tyr Tyr Tyr Tyr Gly Met Asp Val Trp Gly

(2) INFORMATION FOR SEQ ID NO:119:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

Pro Pro

(2) INFORMATION FOR SEQ ID NO:120:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 34 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Thr Lys

(2) INFORMATION FOR SEQ ID NO:121:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

Tyr Cys Ala Arg Arg Gly Val Ser Asp Ala Phe Asp Ile Trp Gly Gln
 1 5 10 15

Gly Thr Met Val Thr Val Ser Ser Ala Asp Thr Lys
 20 25

(2) INFORMATION FOR SEQ ID NO:122:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Tyr Cys Ala Arg Ala Thr Gly Ala Phe Asp Ile Trp Gly Gln Gly Thr
 1 5 10 15

Met Val Thr Val Ser Ser Gly Ser Ala Ser
 20 25

(2) INFORMATION FOR SEQ ID NO:123:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Tyr Cys Ala Arg Ser Ala Asn Trp Gly Ser Tyr Tyr Tyr Tyr Gly Met
 1 5 10 15

Asp Val Trp Gly Gln Gly Thr Thr Val Thr Val Ser Ser Gly Ser Ala
 20 25 30

Ser

(2) INFORMATION FOR SEQ ID NO:124:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

```

Tyr Cys Ala Arg Tyr Phe Gln His Trp Gly Gln Gly Thr Leu Val Thr
1           5           10           15
Val Ser Ser Gly Ser Ala Ser
                20

```

(2) INFORMATION FOR SEQ ID NO:125:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

```

Tyr Cys Ala Arg His Val Ala Asn Ser Phe Asp Tyr Trp Gly Gln Gly
1           5           10           15
Thr Leu Val Thr Val Ser Ser Gly Ser Ala Ser
                20           25

```

(2) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

```

Tyr Cys Ala Arg Gln Ile Thr Met Val Arg Gly Val Pro Phe Asp Tyr
1           5           10           15
Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ser Gly Ser Ala Ser
                20           25           30

```

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

Tyr Cys Ala Arg Gln Tyr Phe Gln His Trp Gly Gln Gly Thr Leu Val
 1 5 10 15

Thr Val Ser Ser Gly Ser Ala Ser
 20

(2) INFORMATION FOR SEQ ID NO:128:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

Tyr Cys Ala Arg Gln Thr Gly Asp Tyr Tyr Tyr Tyr Gly Met Asp Val
 1 5 10 15

Trp Gly Gln Gly Thr Thr Val Thr Val Ser Ser Gly Ser Ala Ser
 20 25 30

(2) INFORMATION FOR SEQ ID NO:129:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 36 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Tyr Cys Ala Arg His Tyr Tyr Gly Ser Gly Ser Tyr Asp Tyr Tyr Tyr
 1 5 10 15

Tyr Gly Met Asp Val Trp Gly Gln Gly Thr Thr Val Thr Val Ser Ser
 20 25 30

Gly Ser Ala Ser
 35

(2) INFORMATION FOR SEQ ID NO:130:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

Tyr	Cys	Ala	Arg	Gln	Gly	Val	Gly	Pro	Gly	Asn	Pro	Gly	His	Arg	Leu
1				5				10						15	
Leu	Ser	Leu	His	Gln											
				20											

(2) INFORMATION FOR SEQ ID NO:131:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Tyr	Cys	Val	Arg	Phe	Trp	Glu	Thr	Gly	Ser	Thr	Pro	Gly	Ala	Arg	Glu
1				5					10					15	
Pro	Trp	Ser	Pro	Ser	Pro	Gln	Gly	Val	His						
				20				25							

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Tyr	Cys	Ala	Arg	Arg	Arg	Tyr	Tyr	Gly	Ser	Gly	Ser	Tyr	Tyr	Asn	Val
1				5					10					15	
Phe	Asp	Tyr	Trp	Gly	Gln	Gly	Thr	Leu	Val	Thr	Val	Ser	Ser	Gly	Ser
			20					25						30	
Thr	Lys														

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

Tyr	Cys	Ala	Arg	Gln	Thr	Trp	Gly	Gly	Asp	Tyr	Trp	Gly	Gln	Gly	Thr
1				5					10					15	
Leu	Val	Thr	Val	Ser	Ser	Gly	Ser	Thr	Lys						
			20					25							

(2) INFORMATION FOR SEQ ID NO:134:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

Tyr	Cys	Ala	Arg	Gly	Tyr	Ser	Gly	Tyr	Asp	Asn	Tyr	Tyr	Tyr	Gly	Ile
1				5					10					15	
His	Val	Trp	Gly	Gln	Gly	Thr	Thr	Val	Thr	Val	Ser	Ser	Gly	Ser	Thr
			20					25					30		
Lys															

(2) INFORMATION FOR SEQ ID NO:135:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

Tyr	Cys	Ala	Arg	Gln	Thr	Gly	Glu	Asp	Tyr	Phe	Asp	Tyr	Trp	Gly	Gln
1				5					10					15	
Gly	Thr	Leu	Val	Thr	Val	Ser	Ser	Gly	Ser	Ala	Ser				
			20					25							

(2) INFORMATION FOR SEQ ID NO:136:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

Tyr	Cys	Ala	Arg	Tyr	Ser	Gly	Tyr	Asp	Tyr	Leu	Leu	Val	Leu	Arg	Ser
1				5				10						15	
Leu	Gly	Pro	Trp	His	Pro	Gly	His	Cys	Leu	Leu	Ser	Leu	His	Arg	
			20					25					30		

(2) INFORMATION FOR SEQ ID NO:137:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Tyr	Cys	Ala	Arg	Ala	Ser	Leu	Pro	Ser	Phe	Asp	Tyr	Tyr	Gly	Met	Asp
1				5				10						15	
Val	Trp	Gly	Gln	Gly	Thr	Thr	Val	Thr	Val	Ser	Ser	Gly	Ser	Thr	Lys
			20					25					30		

(2) INFORMATION FOR SEQ ID NO:138:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

Tyr	Cys	Ala	Arg	Arg	Gly	Gly	Gly	Leu	Thr	Thr	Gly	Ala	Arg	Glu	Pro
1				5				10						15	
Trp	Ser	Pro	Ser	Pro	Gln	Gly	Val	His							
			20					25							

(2) INFORMATION FOR SEQ ID NO:139:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

Tyr	Cys	Val	Pro	Val	Glu	Thr	Leu	Leu	Leu	Leu	Leu	Arg	Tyr	Gly	Arg
1				5				10						15	
Leu	Gly	Pro	Arg	Asp	His	Gly	His	Arg	Leu	Leu	Arg	Glu	Cys	Ile	
			20					25					30		

(2) INFORMATION FOR SEQ ID NO:140:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

Tyr	Cys	Val	Arg	Asp	Ile	Leu	Thr	Gly	Xaa	Arg	Asp	Tyr	Trp	Gly	Gln
1				5				10						15	
Gly	Thr	Leu	Val	Thr	Val	Ser	Ser	Gly	Ser	Ala	Ser				
			20					25							

(2) INFORMATION FOR SEQ ID NO:141:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

Tyr	Cys	Ala	Arg	His	Gly	Ile	Ala	Ala	Ala	Gly	Thr	Ala	Phe	Asp	Ile
1				5				10						15	
Trp	Gly	Gln	Gly	Thr	Met	Val	Thr	Val	Ser	Ser	Gly	Ser	Ala	Ser	
			20					25					30		

(2) INFORMATION FOR SEQ ID NO:142:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 28 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

Tyr	Cys	Val	Arg	Ser	Thr	Gly	Val	Asp	Ala	Phe	Asp	Ile	Trp	Gly	Gln
1				5					10					15	
Gly	Thr	Met	Val	Thr	Val	Ser	Ser	Gly	Ser	Ala	Ser				
			20					25							

(2) INFORMATION FOR SEQ ID NO:143:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

Tyr	Cys	Ala	Glu	Ile	Ala	Ala	Ala	Ala	Leu	Leu	Leu	Xaa	Leu	Leu	Gly
1				5					10					15	
Pro	Gly	Asn	Pro	Gly	His	Arg	Leu	Leu	Arg	Glu	Cys	Ile			
			20					25							

(2) INFORMATION FOR SEQ ID NO:144:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

Tyr	Cys	Val	Cys	Ile	Ala	Ala	Ala	Gly	Lys	Gly	Asn	Gly	Tyr	Trp	Gly
1				5					10					15	
Gln	Gly	Thr	Leu	Val	Thr	Val	Ser	Ser	Gly	Ser	Ala	Ser			
			20					25							

(2) INFORMATION FOR SEQ ID NO:145:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

Tyr	Cys	Ala	Arg	Gln	Asn	Trp	Gly	Asp	Tyr	Trp	Gly	Gln	Gly	Thr	Leu
1				5				10						15	
Val	Thr	Val	Ser	Ser	Gly	Ser	Ala	Ser							
			20					25							

(2) INFORMATION FOR SEQ ID NO:146:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

TCAGTGAAGG TTCCTGCAA GGCATCTGGA TACACCTTCA CC

42

(2) INFORMATION FOR SEQ ID NO:147:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GT

42

(2) INFORMATION FOR SEQ ID NO:148:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 45 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

GGCCGCATCC CGGGTCTCGA GGTGACAAG CTTTCGAGGA TCCGC

45

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

GGCCGCGGAT CCTCGAAAGC TTGTCGACCT CGAGACCCGG GATGC

45

(2) INFORMATION FOR SEQ ID NO:150:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

GGCCGCTGTC GACAAGCTTA TCGATGGATC CTCGAGTGC

39

(2) INFORMATION FOR SEQ ID NO:151:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

GGCCGCACTC GAGGATCCAT CGATAAGCTT GTCGACAGC

39

(2) INFORMATION FOR SEQ ID NO:152:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

CACCTTCGGC CAAGGGACAC GACTGGAGAT TAAACGTAAG CA

42

(2) INFORMATION FOR SEQ ID NO:153:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 48 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

AGGTTTCAGTG GCAGTGGGTC TGGGACAGAC TTCACTCTCA CCATCAGC

48

(2) INFORMATION FOR SEQ ID NO:154:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

GATCCTCGAG ACCAGGTACC AGATCTTGTG AATTCG

36

(2) INFORMATION FOR SEQ ID NO:155:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

TCGACGAATT CACAAGATCT GGTACCTGGT CTCGAG

36

(2) INFORMATION FOR SEQ ID NO:156:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

CGCGGTACCG AGAGTCAGTC CTTCCCAAAT GTC

33

(2) INFORMATION FOR SEQ ID NO:157:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

CGCCTCGAGA CAGCTGGAAT GGGCACATGC AGA

33

(2) INFORMATION FOR SEQ ID NO:158:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

CGCGGTACCG CTGATGCTGC ACCAACTGTA TCC

33

(2) INFORMATION FOR SEQ ID NO:159:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:
CGCCTCGAGC TAACACTCAT TCCTGTTGAA GCT 33
- (2) INFORMATION FOR SEQ ID NO:160:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:
GGCGCTCGAG CTGGACAGGG MTCCAAGTT CCA 33
- (2) INFORMATION FOR SEQ ID NO:161:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:
CCACACTCTG CATGCTGCAG AAGCTTTTCT GTA 33
- (2) INFORMATION FOR SEQ ID NO:162:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:
GGTGACTGAG GTACCTTGAC CCCAGTAGTC CAG 33
- (2) INFORMATION FOR SEQ ID NO:163:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

GGTTACCTCA GTCACCGTCT CCTCAGAGGT AAGAATGGCC TC

42

(2) INFORMATION FOR SEQ ID NO:164:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

AGGCTCCACC AGACCTCTCT AGACAGCAAC TAC

33

(2) INFORMATION FOR SEQ ID NO:165:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 51 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

TGGGGTCAAG GAACCTCAGT CACCGTCTCC TCAGGTAAGA ATGGCCTCTC C

51

(2) INFORMATION FOR SEQ ID NO:166:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

Trp Gly Gln Gly Thr Ser Val Thr Val Ser Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO:167:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 53 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

TGGGGTCAAG GTACCTCAGT CACCGTCTCC TCAGAGGTAA GAATGGCCTC TCC

53

(2) INFORMATION FOR SEQ ID NO:168:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

Trp Gly Gln Gly Thr Ser Val Thr Val Ser Ser Glu
 1 5 10

(2) INFORMATION FOR SEQ ID NO:169:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 46 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

CTGGTCCTCA GAGAGTCAGT CCTTCCCAA TGTCTTCCCC CTCGTC

46

(2) INFORMATION FOR SEQ ID NO:170:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

Glu	Ser	Gln	Ser	Phe	Pro	Asn	Val	Phe	Pro	Leu	Val
1				5					10		

(2) INFORMATION FOR SEQ ID NO:171:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

CTGGTCCTCA GAGTCAGTCC TTCCCGAATG TCTTCCCCCT CGTC

44

(2) INFORMATION FOR SEQ ID NO:172:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

Ser	Gln	Ser	Phe	Pro	Asn	Val	Phe	Pro	Leu	Val
1				5					10	

(2) INFORMATION FOR SEQ ID NO:173:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

TAGAAGGAAT TCAGCAGGCA CACAACAGAG GCAGTTCCA

39

(2) INFORMATION FOR SEQ ID NO:174:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 38 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

AGCTTCTCGA GCTCCTGCTG CTCTGTTTCC CAGGTGCC

38

(2) INFORMATION FOR SEQ ID NO:175:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

CAGCTTCTCG AGCTCCTGCT ACTCTGGCTC MCAGATACC

39

(2) INFORMATION FOR SEQ ID NO:176:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

TATTACTGTG CGAGGGCTCC AACTGGGGAC TGGTTCGAC

39

(2) INFORMATION FOR SEQ ID NO:177:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

Tyr Tyr Cys Ala Arg Ala Pro Thr Gly Asp Trp Phe Asp

1

5

10

(2) INFORMATION FOR SEQ ID NO:178:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

TATAATAGTT ACCCTCCTAC TTTCGGC

27

(2) INFORMATION FOR SEQ ID NO:179:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 9 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

Tyr	Asn	Ser	Tyr	Pro	Pro	Thr	Phe	Gly
1				5				

(2) INFORMATION FOR SEQ ID NO:180:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

GGCGCGCCTT GGCCTAAGAG GCCA

24

(2) INFORMATION FOR SEQ ID NO:181:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

2360

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

CCTCTTAGGC CAAGGCGCGC CTGG

24

(2) INFORMATION FOR SEQ ID NO:182:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

AATTCAGTAT CGATGTGGTA C

21

(2) INFORMATION FOR SEQ ID NO:183:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

CACATCGATA CTG

13

(2) INFORMATION FOR SEQ ID NO:184:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

GTTTGCAGGT GTCCAGTGTS AGGTGCAGCT GKTGGAGTCY SG

42

(2) INFORMATION FOR SEQ ID NO:185:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

CCGGTCGACC CG

12

(2) INFORMATION FOR SEQ ID NO:186:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

TCACAAGCCC AGCAACACCA AG

22

(2) INFORMATION FOR SEQ ID NO:187:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

AAAAGCCAGA AGACCCTCTC CCTG

24

(2) INFORMATION FOR SEQ ID NO:188:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

CAATAGGGGT CATGGACCC

19

(2) INFORMATION FOR SEQ ID NO:189:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

TCATTCTGTG CAGAGTTGGC

20

(2) INFORMATION FOR SEQ ID NO:190:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

GTCCAGAATT CGGTBCAGCT GGTGSAGTCT GG

32

(2) INFORMATION FOR SEQ ID NO:191:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

GGTTTCTCGA GGAAGAGGAA GACTGACGGT CC

32

(2) INFORMATION FOR SEQ ID NO:192:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:
GACATCCAGC TGACCCAGTC TCC 23

(2) INFORMATION FOR SEQ ID NO:193:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:193:
GATATTCAGC TGACTCAGTC TCC 23

(2) INFORMATION FOR SEQ ID NO:194:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:
GAAATTCAGC TGACGCAGTC TCC 23

(2) INFORMATION FOR SEQ ID NO:195:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:
GAAACGCAGC TGACGCAGTC TCC 23

(2) INFORMATION FOR SEQ ID NO:196:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 35 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

GCAAGCTTCT GTCCCAGACC CACTGCCACT GAACC

35

(2) INFORMATION FOR SEQ ID NO:197:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

CGGTTAACAT AGCCCTGGGA CGAGAC

26

(2) INFORMATION FOR SEQ ID NO:198:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

GGGTAACTC ATTGCCTCCA AAGCAC

26

(2) INFORMATION FOR SEQ ID NO:199:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 413 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

TGCACAAGAA CATGAAACAC CTGTGGTTCT TCCTCCTCCT GGTGGCAGCT CCCAGATGGG

60

TCCTGTCCCA GGTGCAGCTT CATCAGTGGG GCGCAGGACT GTTGAAGCCT TCGGAGACCC

120

TGTCCCTCAC CTGCGCTGTC TATGGTGGGT CCTTCAGTGG TTACTTCTGG AGCTGGATCC	180
GCCAGCCCCC AGGGAGGGGG CTGGAGTGGA TTGGGGAAAT CCATCATCGT GGAAGCACCA	240
ACTACAACCC GTCCCTCGAG AGTCGAGTCA CCCTATCAGT AGACACGTCC AAAAACCAGT	300
TCTCCCTGAG GCTGAGTTCT GTGACCGCCG CGGACACGGC TGTGTATTAC TGTGCGAGAG	360
ACATTACTAT GGTTCGGGGA GTACCTCACT GGGGCCAGGG AACCTGGTC ACC	413

(2) INFORMATION FOR SEQ ID NO:200:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 153 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

GACAGACTTC ACTCTACCA TCAGCAGACT GGAGCCTGAA GATTTTGCAG TGTATTACTG	60
TCAGCAGTAT GGTAGCTCAC CCCTCACTTT CGGCGGAGGG ACCAAGGTGG AGATCAAACG	120
AACTGTGGCG GCACCATCTG TCTTCATCTT CCC	153

(2) INFORMATION FOR SEQ ID NO:201:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 414 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

TCCACCATCA TGGGGTCAAC CGCCATCCTC GCCCTCCTCC TGGCTGTTCT CCAAGGAGTC	60
TGTGCCGAGG TGCAGCTGGT GCAGTCTGGA GCAGAGGTGA AAAAGCCCGG GGAGTCTCTG	120
AAGATCTCCT GTAAGGGTTC TGGATACAGC TTTACCACTT ACTGGATCGC CTGGGTGCGC	180
CAGATGCCCC GGAAAGGCCT GGAGTGGATG GGGATCATCG ATCCTGCTGA CTCTGATACC	240
AGATACAACC CGTCCTTCCA AGGCCAGGTC ACCATCTCAG CCGACAAGTC CATCAGTACC	300
GCCTATTTGC AGTGGAGCAG CCTGAAGGCC TCGGACACCG CCATGTATTA CTGTGCGAGA	360
CCAGCGAACT GGAAGTGGTA CTTGTTCTC TGGGGCCGTG GCACCCTGGT CACT	414

(2) INFORMATION FOR SEQ ID NO:202:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 156 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

GACAGATTTT ACTCTACCA TCAGCAGCCT GCAGCCTGAA GATTTTGCAA CTTATTACTG	60
TCAACAGTTT ATTAGTTACC CTCAGCTCAC TTTCGGCGGA GGGACCAGGG TGGAGATCAA	120
ACGAACTGTG GCTGCACCAT CTGTCTTCAT CTTCCC	156

(2) INFORMATION FOR SEQ ID NO:203:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 404 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

TGCACAAGAA CATGAAACAC CTGTGGTTCT TCCTCCTCCT GGTGGCAGCT CCCAGATGGG	60
TCCTGTCCCA GGTGCAGCTA CAGCAGTGGG GCGCAGGACT GTTGAAGCCT TCGGAGACCC	120
TGTCCCTCAC CTGCGCTGTC TATGGTGGGT CCTTCAGTGG TTACTACTGG AGCTGGATCC	180
GCCAGCCCCC AGGTAAGGGG CTGGAGTGGG TTGGGGAAAT CAATCATAGT GGAAGCACCA	240
ACTACAACCC GTCCCTCAAG AGTCGAGTCA CCATATCAGT CGACACGTCC AAGAACCAGT	300
TCTCCCTGAA ACTGAGCTCT GTGACCGCCG CGGACACGGC TGTGTATTAC TGTGCGAGAG	360
TAATTAATTG GTTCGACCCC TGGGGCCAGG GAACCTGGT CACC	404

(2) INFORMATION FOR SEQ ID NO:204:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 153 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:

GACAGATTTC ACTCTCACCA TCAGCAGCCT GCAGCCTGAA GATTTTGCAA CTTACTATTG	60
TCAACAGGCT AATAGTTTCC CGTACACTTT TGGCCAGGGG ACCAAGCTGG AGATCAAACG	120
AACTGTGGCT GCACCATCTG TCTTCATCTT CCC	153

(2) INFORMATION FOR SEQ ID NO:205:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 403 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

ATGAAACACC TGTGGTTCTT CCTCCTCCTG GTGGCAGCTC CCAGATGGGT CCTGTCCCAG	60
GTGCAGCTAC AGCAGTGGGG CGCAGGACTG TTGAAGCCTT CGGAGACCCT GTCCCTCACC	120
TGCGCTGTCT ATGGTGGGTC CTTCACTGGT TACTACTGGA GCTGGATCCG CCAGCCCCCA	180
GGTAAGGGGC TGGAGTGGAT TGGGGAAATC AATCATAGTG GAAGCACCAA CTACAACCCG	240
TCCCTCAAGA GTCGAGTCAC CATATCAGTC GACACGTCCA AGAACCAGTT CTCCTGAAG	300
CTGAGCTCTG TGACCGCCGC GGACACGGCT GTGTATTACT GTGCGAGAGT AATTAATTGG	360
TTGACCCCT GGGGCCAGGG AACCTGGTC ACCGTCTCCT CAG	403

(2) INFORMATION FOR SEQ ID NO:206:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 388 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

ATGGACATGA TGGTCCCCGC TCAGTCCTG GGGCTCCTGC TGCTCTGGTT CCCAGGTTCC	60
AGATGCGACA TCCAGATGAC CCAGTCTCCA TCTTCCGTGT CTGCATCTGT AGGAGACAGA	120
GTCACCATCA CTTGTCTGGC GAGTCAGGAT ATTAGCAGCT GGTAGCCTG GTATCAGCAT	180
AAACCAGGGA AAGCCCCTAA GTCCTGATC TATGCTGCAT CCAGTTTGCA AAGTGGGGTC	240
CCATCAAGGT TCAGCGGCAG TGGATCTGGG ACAGATTTC CTCTCACCAT CAGCAGCCTG	300

CAGCCTGAAG ATTTTGCAAC TTACTATTGT CAACAGGCTA ATAGTTTCCC GTACACTTTT	360
GGCCAGGGGA CCAAGCTGGA GATCAAAC	388

(2) INFORMATION FOR SEQ ID NO:207:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 462 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

ATGGGGTCAA CCGCCATCCT CGCCCTCCTC CTGGCTGTTC TCCAAGGAGT CTGTGCCGAG	60
GTGCAGCTGG TGCAGTCTGG AGCAGAGGTG AAAAAGCCCG GGGAGTCTCT GAAGATCTCC	120
TGTAAGGGTT CTGGATACAG CTTTACCGGC TACTGGATCG GCTGGGTGCG CCAGATGCCC	180
GGGAAAGGCC TGGAGTGGAT GGGGATCATC TATCCTGGTG ACTCTGATAC CACATACAGC	240
CCGTCCTTCC AAGGCCAGGT CACCATCTCA GCCGACAAGT CCATCAGCAC CGCCTACCTG	300
CAGTGGAGCA GCCTGAAGGC CTCGGACACC GCCATGTATT ACTGTGCGAG AGACCAACTG	360
GGCCTCTTTG ACTACTGGGG CCAGGGAACC CTGGTCACCG TCTCCTCAGC CTCCACCAAG	420
GGCCCATCGG TCTTCCCCCT GGCACCCTCC TCCAAGAAGC TT	462

(2) INFORMATION FOR SEQ ID NO:208:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 439 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

ATGGACATGG AGTTCCCCGT TCAGCTCCTG. GGGCTCCTGC TGCTCTGTTT CCCAGGTGCC	60
AGATGTGACA TCCAGATGAC CCAGTCTCCA TCCTCACTGT CTGCATCTGT AGGAGACAGA	120
GTCACCATCA CTTGTCGGGC GAGTCAGGGT ATTAGCAGCT GGTTAGCCTG GTATCAGCAG	180
AAACCAGAGA AAGCCCCTAA GTCCCTGATC TATTCTGCAT CCAGTTTGCA AAGTGGGGTC	240
CCATCAAGGT TCAGCGGCAG TGGATCTGGG ACAGATTTCA CTCTACCAT CAGCAGCCTG	300
CAGCCTGAAG ATTTTGCAAC TTATTACTGC CAACAGTATG ATAGTTACCC GTACACTTTT	360

GGCCAGGGGA CCAAGCTGGA GATCAAACGA ACTGTGGCTG CACCATCTGT CTTTCATCTTC 420

CCGCCATCTG ATGAAGCTT 439

(2) INFORMATION FOR SEQ ID NO:209:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

Asp	Ile	Thr	Met	Val	Arg	Gly	Val	Pro	His
1				5				10	

(2) INFORMATION FOR SEQ ID NO:210:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 9 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

Gln	Gln	Tyr	Gly	Ser	Ser	Pro	Leu	Thr
1				5				

(2) INFORMATION FOR SEQ ID NO:211:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

Pro	Ala	Asn	Trp	Asn	Trp	Tyr	Phe	Val	Leu
1				5				10	

(2) INFORMATION FOR SEQ ID NO:212:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

Gln	Gln	Phe	Ile	Ser	Tyr	Pro	Gln	Leu	Thr
1				5				10	

(2) INFORMATION FOR SEQ ID NO:213:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:

Val	Ile	Asn	Trp	Phe	Asp	Pro
1				5		

(2) INFORMATION FOR SEQ ID NO:214:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:

Gln	Gln	Ala	Asn	Ser	Phe	Pro	Tyr	Thr
1				5				

(2) INFORMATION FOR SEQ ID NO:215:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

Asp	Gln	Leu	Gly	Leu	Phe	Asp	Tyr
1				5			

(2) INFORMATION FOR SEQ ID NO:216:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 9 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

Gln Gln Tyr Asp Ser Tyr Pro Tyr Thr
 1 5

(2) INFORMATION FOR SEQ ID NO:217:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3881 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:

TCTTCCGCTT CCTCGCTCAC TGA	CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA	60
TCAGCTCACT CAAAGGCGGT AATACGGTTA	TCCACAGAAT CAGGGGATAA CGCAGGAAAG	120
AACATGTGAG CAAAAGGCCA GCAA	AAGGCCG GTTGCTGGCG	180
TTTTTCCATA GGCTCCGCC CCCTGACGAG	CATCACAAAA ATCGACGCTC AAGTCAGAGG	240
TGGCGAAACC CGACAGGACT ATAAAGATAC	CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG	300
CGCTCTCCTG TTCCGACCCT GCCGCTTACC	GGATACCTGT CCGCCTTTCT CCCTTCGGGA	360
AGCGTGGCGC TTTCTCATAG CTCACGCTGT	AGGTATCTCA GTTCGGTGTA GGTCGTTGCG	420
TCCAAGCTGG GCTGTGTGCA CGAACCCCC	GTTACAGCCG ACCGCTGCGC CTTATCCGGT	480
AACTATCGTC TTGAGTCCAA CCCGGTAAGA	CACGACTTAT CGCCACTGGC AGCAGCCACT	540
GGTAACAGGA TTAGCAGAGC GAGGTATGTA	GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG	600
CCTAACTACG GCTACACTAG AAGGACAGTA	TTTGGTATCT GCGCTCTGCT GAAGCCAGTT	660
ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA	TCCGGCAAAC AAACCACCGC TGGTAGCGGT	720
GGTTTTTTTG TTTGCAAGCA GCAGATTACG	CGCAGAAAAA AAGGATCTCA AGAAGATCCT	780
TTGATCTTTT CTACGGGGTC TGACGCTCAG	TGGAACGAAA ACTCACGTTA AGGGATTTTG	840

GTCATGAGAT	TATCAAAAAG	GATCTTCACC	TAGATCCTTT	TAAATTAAAA	ATGAAGTTTT	900
AAATCAATCT	AAAGTATATA	TGAGTAAACT	TGGTCTGACA	GTTACCAATG	CTTAATCAGT	960
GAGGCACCTA	TCTCAGCGAT	CTGTCTATTT	CGTTCATCCA	TAGTTGCCTG	ACTCCCCGTC	1020
GTGTAGATAA	CTACGATACG	GGAGGGCTTA	CCATCTGGCC	CCAGTGCTGC	AATGATACCG	1080
CGAGACCCAC	GCTCACCUGC	TCCAGATTTA	TCAGCAATAA	ACCAGCCAGC	CGGAAGGGCC	1140
GAGCGCAGAA	GTGGTCCTGC	AACTTTATCC	GCCTCCATCC	AGTCTATTAA	TTGTTGCCGG	1200
GAAGCTAGAG	TAAGTAGTTC	GCCAGTTAAT	AGTTTGCGCA	ACGTTGTTGC	CATTGCTACA	1260
GGCATCGTGG	TGTCACGCTC	GTCGTTTGGT	ATGGCTTCAT	TCAGCTCCGG	TTCCCAACGA	1320
TCAAGGCGAG	TTACATGATC	CCCCATGTTG	TGCAAAAAAG	CGGTTAGCTC	CTTCGGTCTT	1380
CCGATCGTTG	TCAGAAGTAA	GTTGGCCGCA	GTGTTATCAC	TCATGGTTAT	GGCAGCACTG	1440
CATAATTCTC	TTACTGTCAT	GCCATCCGTA	AGATGCTTTT	CTGTGACTGG	TGAGTACTCA	1500
ACCAAGTCAT	TCTGAGAATA	GTGTATGCGG	CGACCGAGTT	GCTCTTGCCC	GGCGTCAATA	1560
CGGGATAATA	CCGCGCCACA	TAGCAGAACT	TTAAAAGTGC	TCATCATTGG	AAAACGTTCT	1620
TCGGGGCGAA	AACTCTCAAG	GATCTTACCG	CTGTTGAGAT	CCAGTTCGAT	GTAACCCACT	1680
CGTGACCCCA	ACTGATCTTC	AGCATCTTTT	ACTTTCACCA	GCGTTTCTGG	GTGAGCAAAA	1740
ACAGGAAGGC	AAAATGCCGC	AAAAAAGGGA	ATAAGGGCGA	CACGGAAATG	TTGAATACTC	1800
ATACTCTTCC	TTTTTCAATA	TTATTGAAGC	ATTTATCAGG	GTTATTGTCT	CATGAGCGGA	1860
TACATATTTG	AATGTATTTA	GAAAAATAAA	CAAATAGGGG	TTCCGCGCAC	ATTTCCCCGA	1920
AAAGTGCCAC	CTGACGTCTA	AGAAACCATT	ATTATCATGA	CATTAACCTA	TAAAAATAGG	1980
CGTATCACGA	GGCCCTTTTC	TCTCGCGCGT	TTCGGTGATG	ACGGTGAAAA	CCTCTGACAC	2040
ATGCAGCTCC	CGGAGACGGT	CACAGCTTGT	CTGTAAGCGG	ATGCCGGGAG	CAGACAAGCC	2100
CGTCAGGGCG	CGTCAGCGGG	TGTTGGCGGG	TGTCGGGGCT	GGCTTAACTA	TGCGGCATCA	2160
GAGCAGATTG	TACTGAGAGT	GCACCATATG	CGGTGTGAAA	TACCGCACAG	ATGCGTAAGG	2220
AGAAAAATACC	GCATCAGGCG	CCATTCGCCA	TTCAGGCTGC	GCAACTGTTG	GGAAGGGCGA	2280
TCGGTGCGGG	CCTCTTCGCT	ATTACGCCAG	CTGGCGAAAG	GGGGATGTGC	TGCAAGGCGA	2340
TTAAGTTGGG	TAACGCCAGG	GTTTTCCCAG	TCACGACGTT	GTAAAACGAC	GGCCAGTGCC	2400
AAGCTAGCGG	CCGCGGTCCA	ACCACCAATC	TCAAAGCTTG	GTACCCGGGA	GCCTGTTATC	2460
CCAGCACAGT	CCTGGAAGAG	GCACAGGGGA	AATAAAAGCG	GACGGAGGCT	TTCCTTGACT	2520
CAGCCGCTGC	CTGGTCTTCT	TCAGACCTGT	TCTGAATTCT	AAACTCTGAG	GGGGTCGGAT	2580
GACGTGGCCA	TTCTTTGCCT	AAAGCATTGA	GTTTACTGCA	AGGTCAGAAA	AGCATGCAAA	2640
GCCCTCAGAA	TGGCTGCAAA	GAGCTCCAAC	AAAACAATTT	AGAACTTTAT	TAAGGAATAG	2700

GGGGAAGCTA	GGAAGAAACT	CAAAACATCA	AGATTTTAAA	TACGCTTCTT	GGTCTCCTTG	2760
CTATAATTAT	CTGGGATAAG	CATGCTGTTT	TCTGTCTGTC	CCTAACATGC	CCTGTGATTA	2820
TCCGCAAACA	ACACACCCAA	GGGCAGAACT	TTGTTACTTA	AACACCATCC	TGTTTGCTTC	2880
TTTCCTCAGG	AACTGTGGCT	GCACCATCTG	TCTTCATCTT	CCCGCCATCT	GATGAGCAGT	2940
TGAAATCTGG	AACTGCCTCT	GTTGTGTGCC	TGCTGAATAA	CTTCTATCCC	AGAGAGGCCA	3000
AAGTACAGTG	GAAGGTGGAT	AACGCCCTCC	AATCGGGTAA	CTCCCAGGAG	AGTGTACAG	3060
AGCAGGACAG	CAAGGACAGC	ACCTACAGCC	TCAGCAGCAC	CCTGACGCTG	AGCAAAGCAG	3120
ACTACGAGAA	ACACAAAGTC	TACGCCTGCG	AAGTCACCCA	TCAGGGCCTG	AGCTCGCCCCG	3180
TCACAAAGAG	CTTCAACAGG	GGAGAGTGTT	AGAGGGAGAA	GTGCCCCCAC	CTGCTCCTCA	3240
GTTCCAGCCT	GACCCCCTCC	CATCCTTTGG	CCTCTGACCC	TTTTTCCACA	GGGGACCTAC	3300
CCCTATTGCG	GTCCTCCAGC	TCATCTTTCA	CCTCACCCCC	CTCCTCCTCC	TTGGCTTTAA	3360
TTATGCTAAT	GTTGGAGGAG	AATGAATAAA	TAAAGTGAAT	CTTTGCACCT	GTGGTTTCTC	3420
TCTTTCCTCA	ATTTAATAAT	TATTATCTGT	TGTTTACCAA	CTACTCAATT	TCTCTTATAA	3480
GGGACTAAAT	ATGTAGTCAT	CCTAAGGCGC	ATAACCATTT	ATAAAAATCA	TCCTTCATTC	3540
TATTTTACCC	TATCATCCTC	TGCAAGACAG	TCCTCCCTCA	AACCCACAAG	CCTTCTGTCC	3600
TCACAGTCCC	CTGGGCCATG	GATCCTCACA	TCCCAATCCG	CGGCCGCAAT	TCGTAATCAT	3660
GGTCATAGCT	GTTTCCTGTG	TGAAATTGTT	ATCCGCTCAC	AATTCCACAC	AACATACGAG	3720
CCGGAAGCAT	AAAGTGTAAG	GCCTGGGGTG	CCTAATGAGT	GAGCTAACTC	ACATTAATTG	3780
CGTTGCGCTC	ACTGCCCGCT	TTCCAGTCGG	GAAACCTGTC	GTGCCAGCTG	CATTAATGAA	3840
TCGGCCAACG	CGCGGGGAGA	GGCGGTTTGC	GTATTGGGCG	C		3881

(2) INFORMATION FOR SEQ ID NO:218:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4723 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:

GAACTCGAGC	AGCTGAAGCT	TTCTGGGGCA	GGCCAGGCCT	GACCTTGGCT	TTGGGGCAGG	60
GAGGGGGCTA	AGGTGAGGCA	GGTGGCGCCA	GCCAGGTGCA	CACCCAATGC	CCATGAGCCC	120
AGACACTGGA	CGCTGAACCT	CGCGGACAGT	TAAGAACCCA	GGGGCCTCTG	CGCCCTGGGC	180

CCAGCTCTGT	CCCACACCGC	GGTCACATGG	CACCACCTCT	CTTGCAGCCT	CCACCAAGGG	240
CCCATCGGTC	TTCCCCCTGG	CACCCTCCTC	CAAGAGCACC	TCTGGGGGCA	CAGCGGCCCT	300
GGGCTGCCTG	GTCAAGGACT	ACTTCCCCGA	ACCGGTGACG	GTGTCGTGGA	ACTCAGGCGC	360
CCTGACCAGC	GGCGTGCACA	CCTTCCCCGGC	TGTCCTACAG	TCCTCAGGAC	TCTACTCCCT	420
CAGCAGCGTG	GTGACCGTGC	CCTCCAGCAG	CTTGGGCACC	CAGACCTACA	TCTGCAACGT	480
GAATCACAAG	CCCAGCAACA	CCAAGGTGGA	CAAGAAAGTT	GGTGAGAGGC	CAGCACAGGG	540
AGGGAGGGTG	TCTGCTGGAA	GCCAGGCTCA	GCGCTCCTGC	CTGGACGCAT	CCCGGCTATG	600
CAGCCCCAGT	CCAGGGCAGC	AAGGCAGGCC	CCGTCTGCCT	CTTCACCCGG	AGGCCTCTGC	660
CCGCCCCACT	CATGCTCAGG	GAGAGGGTCT	TCTGGCTTTT	TCCCCAGGCT	CTGGGCAGGC	720
ACAGGCTAGG	TGCCCCTAAC	CCAGGCCCTG	CACACAAAGG	GGCAGGTGCT	GGGCTCAGAC	780
CTGCCAAGAG	CCATATCCGG	GAGGACCCTG	CCCCTGACCT	AAGCCCACCC	CAAAGGCCAA	840
ACTCTCCACT	CCCTCAGCTC	GGACACCTTC	TCTCCTCCCA	GATTCCAGTA	ACTCCCAATC	900
TTCTCTCTGC	AGAGCCCAAA	TCTTGTGACA	AAACTCACAC	ATGCCACCG	TGCCCAGGTA	960
AGCCAGCCCA	GGCCTCGCCC	TCCAGCTCAA	GGCGGGACAG	GTGCCCTAGA	GTAGCCTGCA	1020
TCCAGGGACA	GGCCCCAGCC	GGGTGCTGAC	ACGTCCACCT	CCATCTCTTC	CTCAGCACCT	1080
GAACTCCTGG	GGGGACCGTC	AGTCTTCCTC	TTCCCCCCTAA	AACCCAAGGA	CACCCTCATG	1140
ATCTCCCGGA	CCCCTGAGGT	CACATGCGTG	GTGGTGGACG	TGAGCCACGA	AGACCCTGAG	1200
GTCAAGTTCA	ACTGGTACGT	GGACGGCGTG	GAGGTGCATA	ATGCCAAGAC	AAAGCCGCGG	1260
GAGGAGCAGT	ACAACAGCAC	GTACCGTGTG	GTCAGCGTCC	TCACCGTCCT	GCACCAGGAC	1320
TGGCTGAATG	GCAAGGAGTA	CAAGTGCAAG	GTCTCCAACA	AAGCCCTCCC	AGCCCCCATC	1380
GAGAAAACCA	TCTCCAAAGC	CAAAGGTGGG	ACCCGTGGGG	TGCGAGGGCC	ACATGGACAG	1440
AGGCCGGCTC	GGCCCACCCT	CTGCCCTGAG	AGTGACCGCT	GTACCAACCT	CTGTCCCTAC	1500
AGGGCAGCCC	CGAGAACCAC	AGGTGTACAC	CCTGCCCCCA	TCCCGGGATG	AGCTGACCAA	1560
GAACCAGGTC	AGCCTGACCT	GCCTGGTCAA	AGGCTTCTAT	CCCAGCGACA	TCGCCGTGGA	1620
GTGGGAGAGC	AATGGGCAGC	CGGAGAACAA	CTACAAGACC	ACGCCTCCCG	TGCTGGACTC	1680
CGACGGCTCC	TTCTTCCTCT	ACAGCAAGCT	CACCGTGGAC	AAGAGCAGGT	GGCAGCAGGG	1740
GAACGTCTTC	TCATGCTCCG	TGATGCATGA	GGCTCTGCAC	AACCACTACA	CGCAGAAGAG	1800
CCTCTCCCTG	TCTCCGGGTA	AATGAGTGCG	ACGGCCGGCA	AGCCCCCGCT	CCCCGGGCTC	1860
TCGCGGTGCG	ACGAGGATGC	TTGGCACGTA	CCCCCTGTAC	ATACTTCCCG	GGCGCCCAGC	1920
ATGGAAATAA	AGCACCCAGC	GCTGCCCTGG	GCCCCGTGCGA	GACTGTGATG	GTTCTTTCCA	1980
CGGGTCAGGC	CGAGTCTGAG	GCCTGAGTGG	CATGAGGGAG	GCAGAGCGGG	TCCCACTGTC	2040

CCCACACTGG	CCCAGGCTGT	GCAGGTGTGC	CTGGGCCCCC	TAGGGTGGGG	CTCAGCCAGG	2100
GGCTGCCCTC	GGCAGGGTGG	GGGATTTGCC	AGCGTGGCCC	TCCCTCCAGC	AGCACCTGCC	2160
CTGGGCTGGG	CCACGGGAAG	CCCTAGGAGC	CCCTGGGGAC	AGACACACAG	CCCCTGCCTC	2220
TGTAGGAGAC	TGTCCTGTTC	TGTGAGCGCC	CCTGTCCTCC	CGACCTCCAT	GCCCACTCGG	2280
GGGCATGCCT	GCAGGTCGAC	TCTAGAGGAT	CCCCGGGTAC	CGAGCTCGAA	TTCATCGATG	2340
ATATCAGATC	TGCCGGTCTC	CCTATAGTGA	GTCGTATTAA	TTTCGATAAG	CCAGGTTAAC	2400
CTGCATTAAT	GAATCGGCCA	ACGCGCGGGG	AGAGGCGGTT	TGCGTATTGG	GCGCTCTTCC	2460
GCTTCCTCGC	TCCTGACTC	GCTGCGCTCG	GTCGTTCGGC	TGCGGCGAGC	GGTATCAGCT	2520
CACTCAAAGG	CGGTAATACG	GTTATCCACA	GAATCAGGGG	ATAACGCAGG	AAAGAACATG	2580
TGAGCAAAAG	GCCAGCAAAA	GGCCAGGAAC	CGTAAAAAGG	CCGCGTTGCT	GGCGTTTTTC	2640
CATAGGCTCC	GCCCCCCTGA	CGAGCATCAC	AAAAATCGAC	GCTCAAGTCA	GAGGTGGCGA	2700
AACCCGACAG	GACTATAAAG	ATACCAGGCG	TTTCCCCCTG	GAAGCTCCCT	CGTGCGCTCT	2760
CCTGTTCCGA	CCCTGCCGCT	TACCGGATAC	CTGTCCGCCT	TTCTCCCTTC	GGGAAGCGTG	2820
GCGCTTTCTC	AATGCTCACG	CTGTAGGTAT	CTCAGTTCGG	TGTAGGTCGT	TCGCTCCAAG	2880
CTGGGCTGTG	TGCACGAACC	CCCCGTTCAG	CCCGACCGCT	GCGCCTTATC	CGGTAACATAT	2940
CGTCTTGAGT	CCAACCCGGT	AAGACACGAC	TTATCGCCAC	TGGCAGCAGC	CACTGGTAAC	3000
AGGATTAGCA	GAGCGAGGTA	TGTAGGCGGT	GCTACAGAGT	TCTTGAAGTG	GTGGCCTAAC	3060
TACGGCTACA	CTAGAAGGAC	AGTATTTGGT	ATCTGCGCTC	TGCTGAAGCC	AGTTACCTTC	3120
GGAAAAAGAG	TTGGTAGCTC	TTGATCCGGC	AAACAAACCA	CCGCTGGTAG	CGGTGGTTTTT	3180
TTTGTTTGCA	AGCAGCAGAT	TACGCGCAGA	AAAAAAGGAT	CTCAAGAAGA	TCCTTTGATC	3240
TTTTCTACGG	GGTCTGACGC	TCAGTGGAAC	GAAAACTCAC	GTTAAGGGAT	TTTGGTCATG	3300
AGATTATCAA	AAAGGATCTT	CACCTAGATC	CTTTTAAATT	AAAAATGAAG	TTTTAAATCA	3360
ATCTAAAGTA	TATATGAGTA	AACTTGGTCT	GACAGTTACC	AATGCTTAAT	CAGTGAGGCA	3420
CCTATCTCAG	CGATCTGTCT	ATTTTCGTTCA	TCCATAGTTG	CCTGACTCCC	CGTCGTGTAG	3480
ATAACTACGA	TACGGGAGGG	CTTACCATCT	GGCCCCAGTG	CTGCAATGAT	ACCGCGAGAC	3540
CCACGCTCAC	CGGCTCCAGA	TTTATCAGCA	ATAAACCAGC	CAGCCGGAAG	GGCCGAGCGC	3600
AGAAGTGGTC	CTGCAACTTT	ATCCGCCTCC	ATCCAGTCTA	TTAATTGTTG	CCGGGAAGCT	3660
AGAGTAAGTA	GTTCCGCAGT	TAATAGTTTG	CGCAACGTTG	TTGCCATTGC	TACAGGCATC	3720
GTGGTGTCAC	GCTCGTCGTT	TGGTATGGCT	TCATTGAGCT	CCGGTTCCCA	ACGATCAAGG	3780
CGAGTTACAT	GATCCCCCAT	GTTGTGCAAA	AAAGCGGTTA	GCTCCTTCGG	TCCTCCGATC	3840

TTTGTTCAGAA	GTAAGTTGGC	CGCAGTGTTA	TCACTCATGG	TTATGGCAGC	ACTGCATAAT	3900
TCTCTTACTG	TCATGCCATC	CGTAAGATGC	TTTTCTGTGA	CTGGTGAGTA	CTCAACCAAG	3960
TCATTCTGAG	AATAGTGTAT	GCGGCGACCG	AGTTGCTCTT	GCCCGGCGTC	AATACGGGAT	4020
AATACCGCGC	CACATAGCAG	AACTTTAAAA	GTGCTCATCA	TTGGAAAACG	TTCTTCGGGG	4080
CGAAAACCTCT	CAAGGATCTT	ACCGCTGTTG	AGATCCAGTT	CGATGTAACC	CACTCGTGCA	4140
CCCAACTGAT	CTTCAGCATC	TTTTACTTTC	ACCAGCGTTT	CTGGGTGAGC	AAAAACAGGA	4200
AGGCAAAAATG	CCGCAAAAAA	GGAATAAGG	GCGACACGGA	AATGTTGAAT	ACTCATACTC	4260
TTCTTTTTTC	AATATTATTG	AAGCATTTAT	CAGGGTTATT	GTCTCATGAG	CGGATACATA	4320
TTTGAATGTA	TTTAGAAAAA	TAAACAAATA	GGGGTTCCGC	GCACATTTCC	CCGAAAAGTG	4380
CCACCTGACG	TCTAAGAAAC	CATTATTATC	ATGACATTAA	CCTATAAAAA	TAGGCGTATC	4440
ACGAGGCCCT	TTCGTCTCGC	GCGTTTCGGT	GATGACGGTG	AAAACCTCTG	ACACATGCAG	4500
CTCCCGGAGA	CGGTCACAGC	TTGTCTGTAA	GCGGATGCCG	GGAGCAGACA	AGCCCGTCAG	4560
GGCGCGTCAG	CGGGTGTTGG	CGGGTGTCGG	GGCTGGCTTA	ACTATGCGGC	ATCAGAGCAG	4620
ATTGTACTGA	GAGTGCACCA	TATGGACATA	TTGTCGTTAG	AACGCGGCTA	CAATTAATAC	4680
ATAACCTTAT	GTATCATACA	CATACGATTT	AGGTGACACT	ATA		4723

(2) INFORMATION FOR SEQ ID NO:219:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 524 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:

AAGCTTGCCA	CCATGAAACA	CCTGTGGTTC	TTCTCCTCC	TGGTGGCAGC	TCCTAGATGG	60
GTCTGTCTC	AGGTGCAGCT	ACAGCAGTGG	GGCGCAGGAC	TGTTGAAGCC	TTCGGAGACC	120
CTGTCCCTCA	CCTGCGCTGT	CTATGGTGGT	TCCTTCAGTG	GTTACTACTG	GAGCTGGATC	180
CGCCAGCCAC	CAGGTAAGGG	TCTGGAGTGG	ATTGGTGAAA	TCAATCATAG	TGGAAGCACC	240
AACTACAACC	CGTCTCTCAA	GAGTCGAGTC	ACCATATCAG	TAGACACGTC	CAAGAACCAG	300
TTCTCTCTGA	AACTGAGCTC	TGTGACCGCT	GCGGACACGG	CTGTGTATTA	CTGTGCGAGA	360
GTAATTAATT	GGTTCGACCC	TTGGGGCCAG	GGAACCCTGG	TCACCGTCTC	CTCAGCCTCA	420
ACCAAGGGCC	CATCGGTCTT	CCCCCTGGCA	CCCTCCTCCA	AGAGCACCTC	TGGGGGCACA	480
GCGGCCCTGG	GCTGCCTGGT	CAAGGACTAC	TTCCCCGAAC	CGGT		524

(2) INFORMATION FOR SEQ ID NO:220:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 420 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:

```

AAGCTTGCCA CCATGATGGT CCCAGCTCAG CTCCTCGGTC TCCTGCTGCT CTGGTTCCCA      60
GGTTCCAGAT GCGACATCCA GATGACCCAG TCTCCATCTT CCGTGTCTGC ATCTGTAGGA      120
GACAGAGTCA CCATCACTTG TCGGGCGAGT CAGGATATTA GCAGCTGGTT AGCCTGGTAT      180
CAGCATAAAC CAGGTAAAGC ACCTAAGCTC CTGATCTATG CTGCATCCAG TTTGCAAAGT      240
GGTGTCCCAT CAAGGTTTCTAG CGGAAGTGGA TCTGGGACAG ATTTCACTCT CACCATCAGC      300
AGCCTGCAGC CTGAAGATTT TGCAACTTAC TATTGTCAAC AGGCTAATAG TTTCCCGTAC      360
ACTTTTGGTC AGGGAACCAA GCTGGAGATC AAACGAACTG TGGCTGCACC ATCTGTCTTC      420

```

(2) INFORMATION FOR SEQ ID NO:221:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

```

ATGGTCCCAG CTCAGCTCCT CGGTCTCCTG CTGCTCTGGT TCCC      44

```

(2) INFORMATION FOR SEQ ID NO:222:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

AGGTTCCAGA TGCGACATCC AGATGACCCA GTCTCCATCT TCCG

44

(2) INFORMATION FOR SEQ ID NO:223:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

TGTCTGCATC TGTAGGAGAC AGAGTCACCA TCACTTGTCTG GGCG

44

(2) INFORMATION FOR SEQ ID NO:224:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

AGTCAGGATA TTAGCAGCTG GTTAGCCTGG TATCAGCATA AACC

44

(2) INFORMATION FOR SEQ ID NO:225:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

AGGTAAAGCA CCTAAGCTCC TGATCTATGC TGCATCCAGT TTGC

44

(2) INFORMATION FOR SEQ ID NO:226:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

AGGAGCTTAG GTGCTTTACC TGGTTTATGC TGATACCAGG CTAA

44

(2) INFORMATION FOR SEQ ID NO:227:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

CCAGCTGCTA ATATCCTGAC TCGCCCGACA AGTGATGGTG ACTC

44

(2) INFORMATION FOR SEQ ID NO:228:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

TGTCTCCTAC AGATGCAGAC ACGGAAGATG GAGACTGGGT CATC

44

(2) INFORMATION FOR SEQ ID NO:229:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

TGGATGTCGC ATCTGGAACC TGGGAACCAG AGCAGCAGGA GACC

44

(2) INFORMATION FOR SEQ ID NO:230:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

GAGGAGCTGA GCTGGGACCA TCATGGTGGC AAGCTTAGAG TC

42

(2) INFORMATION FOR SEQ ID NO:231:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:

GACTCTAAGC TTGCCACCAT GATGGTCC

28

(2) INFORMATION FOR SEQ ID NO:232:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

ACCTTGATGG GACACCACTT TGCAAAGTGG ATGCAGCATA GATC

44

(2) INFORMATION FOR SEQ ID NO:233:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:

AAAGTGGTGT CCCATCAAGG TTCAGCGGAA GTGGATCTGG GACA

44

(2) INFORMATION FOR SEQ ID NO:234:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

GATTTCACTC TCACCATCAG CAGCCTGCAG CCTGAAGATT TTGC

44

(2) INFORMATION FOR SEQ ID NO:235:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

AACTTACTAT TGTCAACAGG CTAATAGTTT CCCGTACACT TTTG

44

(2) INFORMATION FOR SEQ ID NO:236:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:236:

GTCAGGGAAC CAAGCTGGAG ATCAAACGAA CTGTGGCTGC ACCA

44

(2) INFORMATION FOR SEQ ID NO:237:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

TCTGTCTTCA TCTTCCCGCC ATCTGATGAG CAGTTGA

37

(2) INFORMATION FOR SEQ ID NO:238:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 40 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

GGGAAGATGA AGACAGATGG TGCAGCCACA GTTCGTTTGA

40

(2) INFORMATION FOR SEQ ID NO:239:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

TCTCCAGCTT GGTTCCTGA CCAAAAGTGT ACGGGAAACT ATTA

44

(2) INFORMATION FOR SEQ ID NO:240:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

GCCTGTTGAC AATAGTAAGT TGCAAAATCT TCAGGCTGCA GGCT

44

(2) INFORMATION FOR SEQ ID NO:241:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

GCTGATGGTG AGAGTGAAAT CTGTCCCAGA TCCACTTCCG CTGA

44

(2) INFORMATION FOR SEQ ID NO:242:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

TCAACTGCTC ATCAGATGGC

20

(2) INFORMATION FOR SEQ ID NO:243:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3819 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:

TCTTCCGCTT CCTCGCTCAC TGA	60
CTCGCTGCG GCGAGCGGTA	
TCAGCTCACT CAAAGGCGGT AATACGGTTA TCCACAGAAT CAGGGGATAA CGCAGGAAAG	120
AACATGTGAG CAAAAGGCCA GCAAAAGGCC AGGAACCGTA AAAAGGCCGC GTTGCTGGCG	180
TTTTTCCATA GGCTCCGCCC CCCTGACGAG CATCACAAAA ATCGACGCTC AAGTCAGAGG	240
TGGCGAAACC CGACAGGACT ATAAAGATAC CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG	300
CGCTCTCCTG TTCCGACCCT GCCGCTTACC GGATACCTGT CCGCCTTTCT CCCTTCGGGA	360
AGCGTGGCGC TTTCTCATAG CTCACGCTGT AGGTATCTCA GTTCGGTGTA GGTCGTTTCG	420
TCCAAGCTGG GCTGTGTGCA CGAACCCCCC GTTCAGCCCG ACCGCTGCGC CTTATCCGGT	480
AACTATCGTC TTGAGTCCAA CCCGGTAAGA CACGACTTAT CGCCACTGGC AGCAGCCACT	540
GGTAACAGGA TTAGCAGAGC GAGGTATGTA GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG	600

CCTAACTACG	GCTACACTAG	AAGGACAGTA	TTTGGTATCT	GCGCTCTGCT	GAAGCCAGTT	660
ACCTTCGGAA	AAAGAGTTGG	TAGCTCTTGA	TCCGGCAAAC	AAACCACCGC	TGGTAGCGGT	720
GGTTTTTTTG	TTTGCAAGCA	GCAGATTACG	CGCAGAAAAA	AAGGATCTCA	AGAAGATCCT	780
TTGATCTTTT	CTACGGGGTC	TGACGCTCAG	TGGAACGAAA	ACTCACGTTA	AGGGATTTTG	840
GTCATGAGAT	TATCAAAAAG	GATCTTCACC	TAGATCCTTT	TAAATTAAAA	ATGAAGTTTT	900
AAATCAATCT	AAAGTATATA	TGAGTAAACT	TGGTCTGACA	GTTACCAATG	CTTAATCAGT	960
GAGGCACCTA	TCTCAGCGAT	CTGTCTATTT	CGTTCATCCA	TAGTTGCCTG	ACTCCCCGTC	1020
GTGTAGATAA	CTACGATACG	GGAGGGCTTA	CCATCTGGCC	CCAGTGCTGC	AATGATACCG	1080
CGAGACCCAC	GCTCACCGGC	TCCAGATTTA	TCAGCAATAA	ACCAGCCAGC	CGGAAGGGCC	1140
GAGCGCAGAA	GTGGTCCTGC	AACTTTATCC	GCCTCCATCC	AGTCTATTAA	TTGTTGCCGG	1200
GAAGCTAGAG	TAAGTAGTTC	GCCAGTTAAT	AGTTTGCGCA	ACGTTGTTGC	CATTGCTACA	1260
GGCATCGTGG	TGTCACGCTC	GTCGTTTGGT	ATGGCTTCAT	TCAGCTCCGG	TTCCCAACGA	1320
TCAAGGCGAG	TTACATGATC	CCCCATGTTG	TGCAAAAAAG	CGGTTAGCTC	CTTCGGTCTT	1380
CCGATCGTTG	TCAGAAGTAA	GTTGGCCGCA	GTGTTATCAC	TCATGGTTAT	GGCAGCACTG	1440
CATAATTCTC	TTACTGTCAT	GCCATCCGTA	AGATGCTTTT	CTGTGACTGG	TGAGTACTCA	1500
ACCAAGTCAT	TCTGAGAATA	GTGTATGCGG	CGACCGAGTT	GCTCTTGCCC	GGCGTCAATA	1560
CGGGATAATA	CCGCGCCACA	TAGCAGAACT	TTAAAAGTGC	TCATCATTGG	AAAACGTTCT	1620
TCGGGGCGAA	AACTCTCAAG	GATCTTACCG	CTGTTGAGAT	CCAGTTCGAT	GTAACCCACT	1680
CGTGCACCCA	ACTGATCTTC	AGCATCTTTT	ACTTTCACCA	GCGTTTCTGG	GTGAGCAAAA	1740
ACAGGAAGGC	AAAATGCCGC	AAAAAAGGGA	ATAAGGGCGA	CACGGAAATG	TTGAATACTC	1800
ATACTCTTCC	TTTTTCAATA	TTATTGAAGC	ATTTATCAGG	GTTATTGTCT	CATGAGCGGA	1860
TACATATTTG	AATGTATTTA	GAAAAATAAA	CAAATAGGGG	TTCCGCGCAC	ATTTCCCCGA	1920
AAAGTGCCAC	CTGACGTCTA	AGAAACCATT	ATTATCATGA	CATTAACCTA	TAAAAATAGG	1980
CGTATCACGA	GGCCCTTTTCG	TCTCGCGCGT	TTCGGTGATG	ACGGTGAAAA	CCTCTGACAC	2040
ATGCAGCTCC	CGGAGACGGT	CACAGCTTGT	CTGTAAGCGG	ATGCCGGGAG	CAGACAAGCC	2100
CGTCAGGGCG	CGTCAGCGGG	TGTTGGCGGG	TGTCGGGGCT	GGCTTAACTA	TGCGGCATCA	2160
GAGCAGATTG	TACTGAGAGT	GCACCATATG	CGGTGTGAAA	TACCGCACAG	ATGCGTAAGG	2220
AGAAAAATACC	GCATCAGGCG	CCATTCGCCA	TTCAGGCTGC	GCAACTGTTG	GGAAGGGCGA	2280
TCGGTGCGGG	CCTCTTCGCT	ATTACGCCAG	CTGGCGAAAG	GGGGATGTGC	TGCAAGGCGA	2340
TTAAGTTGGG	TAACGCCAGG	GTTTTCCCAG	TCACGACGTT	GTAAAACGAC	GGCCAGTGCC	2400

AAGCTAGCGG	CCGCGGTCCA	ACCACCAATC	TCAAAGCTTG	CCACCATGAT	GGTCCCAGCT	2460
CAGCTCCTCG	GTCTCCTGCT	GCTCTGGTTC	CCAGGTTCCA	GATGCGACAT	CCAGATGACC	2520
CAGTCTCCAT	CTTCCGTGTC	TGCATCTGTA	GGAGACAGAG	TCACCATCAC	TTGTCTGGGCG	2580
AGTCAGGATA	TTAGCAGCTG	GTTAGCCTGG	TATCAGCATA	AACCAGGTAA	AGCACCTAAG	2640
CTCCTGATCT	ATGCTGCATC	CAGTTTGCAA	AGTGGTGTCC	CATCAAGGTT	CAGCGGAAGT	2700
GGATCTGGGA	CAGATTTTAC	TCTCACCATC	AGCAGCCTGC	AGCCTGAAGA	TTTTGCAACT	2760
TACTATTGTC	AACAGGCTAA	TAGTTTCCCG	TACACTTTTG	GTCAGGGAAC	CAAGCTGGAG	2820
ATCAAACGAA	CTGTGGCTGC	ACCATCTGTC	TTCATCTTCC	CGCCATCTGA	TGAGCAGTTG	2880
AAATCTGGAA	CTGCCTCTGT	TGTGTGCCTG	CTGAATAACT	TCTATCCCAG	AGAGGCCAAA	2940
GTACAGTGGA	AGGTGGATAA	CGCCCTCCAA	TCGGGTA ACT	CCCAGGAGAG	TGTCACAGAG	3000
CAGGACAGCA	AGGACAGCAC	CTACAGCCTC	AGCAGCACCC	TGACGCTGAG	CAAAGCAGAC	3060
TACGAGAAAC	ACAAAGTCTA	CGCCTGCGAA	GTCACCCATC	AGGGCCTGAG	CTCGCCCGTC	3120
ACAAAGAGCT	TCAACAGGGG	AGAGTGTTAG	AGGGAGAAGT	GCCCCACCT	GCTCCTCAGT	3180
TCCAGCCTGA	CCCCCTCCCA	TCCTTTGGCC	TCTGACCCTT	TTCCACAGG	GGACCTACCC	3240
CTATTGCGGT	CCTCCAGCTC	ATCTTTCACC	TCACCCCCCT	CCTCCTCCTT	GGCTTTAATT	3300
ATGCTAATGT	TGGAGGAGAA	TGAATAAATA	AAGTGAATCT	TTGCACCTGT	GGTTTCTCTC	3360
TTTCCTCAAT	TTAATAATTA	TTATCTGTTG	TTTACCAACT	ACTCAATTTC	TCTTATAAGG	3420
GACTAAATAT	GTAGTCATCC	TAAGGCGCAT	AACCATTTAT	AAAAATCATC	CTTCATTCTA	3480
TTTTACCCTA	TCATCCTCTG	CAAGACAGTC	CTCCCTCAAA	CCCACAAGCC	TTCTGTCCTC	3540
ACAGTCCCCT	GGGCCATGGA	TCCTCACATC	CCAATCCGCG	GCCGCAATTC	GTAATCATGG	3600
TCATAGCTGT	TTCCTGTGTG	AAATTGTTAT	CCGCTCACAA	TTCCACACAA	CATACGAGCC	3660
GGAAGCATAA	AGTGTAAGC	CTGGGGTGCC	TAATGAGTGA	GCTAACTCAC	ATTAATTGCG	3720
TTGCGCTCAC	TGCCCCGCTT	CCAGTCGGGA	AACCTGTCGT	GCCAGCTGCA	TTAATGAATC	3780
GGCCAACGCG	CGGGGAGAGG	CGGTTTTCGT	ATTGGGCGC			3819

(2) INFORMATION FOR SEQ ID NO:244:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:

362

TTCTTCCTCC TCCTGGTGGC AGCTCCTAGA TGGGTCCTGT CTC

43

(2) INFORMATION FOR SEQ ID NO:245:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

AGGTGCAGCT ACAGCAGTGG GGCGCAGGAC TGTGAAGCC TTC

43

(2) INFORMATION FOR SEQ ID NO:246:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

GGAGACCCTG TCCCTCACCT GCGCTGTCTA TGGTGGTTCC TTC

43

(2) INFORMATION FOR SEQ ID NO:247:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

AGTGGTTACT ACTGGAGCTG GATCCGCCAG CCACCAGGTA AGG

43

(2) INFORMATION FOR SEQ ID NO:248:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

363

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

GTCTGGAGTG GATTGGTGAA ATCAATCATA GTGGAAGCAC CAA

43

(2) INFORMATION FOR SEQ ID NO:249:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

TTCACCAATC CACTCCAGAC CCTTACCTGG TGGCTGGCGG ATC

43

(2) INFORMATION FOR SEQ ID NO:250:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

CAGCTCCAGT AGTAACCACT GAAGGAACCA CCATAGACAG CGC

43

(2) INFORMATION FOR SEQ ID NO:251:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

AGGTGAGGGA CAGGGTCTCC GAAGGCTTCA ACAGTCCTGC GCC

43

(2) INFORMATION FOR SEQ ID NO:252:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

CCACTGCTGT AGCTGCACCT GAGACAGGAC CCATCTAGGA GCT

43

(2) INFORMATION FOR SEQ ID NO:253:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 48 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

GCCACCAGGA GGAGGAAGAA CCACAGGTGT TTCATGGTGG CAAGCTTG

48

(2) INFORMATION FOR SEQ ID NO:254:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

CATGAAACAC CTGTGGTTCT TCC

23

(2) INFORMATION FOR SEQ ID NO:255:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:

TCTTGAGAGA CGGGTTGTAG TTGGTGCTTC CACTATGATT GAT

43

(2) INFORMATION FOR SEQ ID NO:256:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

CTACAACCCG TCTCTCAAGA GTCGAGTCAC CATATCAGTA GAC

43

(2) INFORMATION FOR SEQ ID NO:257:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

ACGTCCAAGA ACCAGTTCTC TCTGAAACTG AGCTCTGTGA CCG

43

(2) INFORMATION FOR SEQ ID NO:258:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

CTGCGGACAC GGCTGTGTAT TACTGTGCGA GAGTAATTAA TTG

43

(2) INFORMATION FOR SEQ ID NO:259:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

366

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:
GTTCGACCCT TGGGGCCAGG GAACCCTGGT CACCGTCTCC TCA 43
- (2) INFORMATION FOR SEQ ID NO:260:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:
GCCTCAACCA AGGGCCCATC GGTCTTCCCC CTGGCACC 38
- (2) INFORMATION FOR SEQ ID NO:261:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:
CGATGGGCCC TTGGTTGAGG CTGAGGAGAC GGTGACCAGG GTTC 44
- (2) INFORMATION FOR SEQ ID NO:262:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:
CCTGGCCCCA AGGGTCGAAC CAATTAATTA CTCTCGCACA GTA 43
- (2) INFORMATION FOR SEQ ID NO:263:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

ATACACAGCC GTGTCCGCAG CGGTCACAGA GCTCAGTTTC AGA

43

(2) INFORMATION FOR SEQ ID NO:264:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 43 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:264:

GAGAACTGGT TCTTGGACGT GTCTACTGAT ATGGTGACTC GAC

43

(2) INFORMATION FOR SEQ ID NO:265:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:265:

GAAGCACCAA CTACAACCCG

20

(2) INFORMATION FOR SEQ ID NO:266:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:

GAGTTCCACG ACACCGTCAC C

21

(2) INFORMATION FOR SEQ ID NO:267:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:

GACCTCAAGC TTGCCACCAT GAAACACCTG TGG

33

(2) INFORMATION FOR SEQ ID NO:268:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4926 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:

GAACTCGAGC AGCTGAAGCT TGCCACCATG AAACACCTGT GGTTCCTCCT CCTCCTGGTG	60
GCAGCTCCTA GATGGGTCCT GTCTCAGGTG CAGCTACAGC AGTGGGGCGC AGGACTGTTG	120
AAGCCTTCGG AGACCCTGTC CCTCACCTGC GCTGTCTATG GTGGTTCCTT CAGTGGTTAC	180
TACTGGAGCT GGATCCGCCA GCCACCAGGT AAGGGTCTGG AGTGGATTGG TGAAATCAAT	240
CATAGTGGAA GCACCAACTA CAACCCGTCT CTCAAGAGTC GAGTCACCAT ATCAGTAGAC	300
ACGTCCAAGA ACCAGTTCTC TCTGAAACTG AGCTCTGTGA CCGCTGCGGA CACGGCTGTG	360
TATTACTGTG CGAGAGTAAT TAATTGGTTC GACCCTTGGG GCCAGGGAAC CCTGGTCACC	420
GTCTCCTCAG CCTCAACCAA GGGCCCATCG GTCTTCCCC TGGCACCTC CTCCAAGAGC	480
ACCTCTGGGG GCACAGCGGC CCTGGGCTGC CTGGTCAAGG ACTACTTCCC CGAACCGGTG	540
ACGGTGTCGT GGAACTCAGG CGCCCTGACC AGCGGCGTGC ACACCTTCCC GGCTGTCCTA	600
CAGTCCTCAG GACTCTACTC CCTCAGCAGC GTGGTGACCG TGCCCTCCAG CAGCTTGGGC	660
ACCCAGACCT ACATCTGCAA CGTGAATCAC AAGCCCAGCA ACACCAAGGT GGACAAGAAA	720
GTGGGTGAGA GGCCAGCACA GGGAGGGAGG GTGTCTGCTG GAAGCCAGGC TCAGCGCTCC	780
TGCCTGGACG CATCCCGGCT ATGCAGCCCC AGTCCAGGGC AGCAAGGCAG GCCCCGTCTG	840
CCTCTTCACC CGGAGGCCTC TGCCCCCCCC ACTCATGCTC AGGGAGAGGG TCTTCTGGCT	900

TTTTCCCCAG	GCTCTGGGCA	GGCACAGGCT	AGGTGCCCCCT	AACCCAGGCC	CTGCACACAA	960
AGGGGCAGGT	GCTGGGCTCA	GACCTGCCAA	GAGCCATATC	CGGGAGGACC	CTGCCCCCTGA	1020
CCTAAGCCCA	CCCCAAAGGC	CAAACCTCTCC	ACTCCCTCAG	CTCGGACACC	TTCTCTCCTC	1080
CCAGATTCCA	GTAACCTCCA	ATCTTCTCTC	TGCAGAGCCC	AAATCTTGTG	ACAAAACCTCA	1140
CACATGCCCA	CCGTGCCCAG	GTAAGCCAGC	CCAGGCCTCG	CCCTCCAGCT	CAAGGCGGGA	1200
CAGGTGCCCT	AGAGTAGCCT	GCATCCAGGG	ACAGGCCCCCA	GCCGGGTGCT	GACACGTCCA	1260
CCTCCATCTC	TTCCTCAGCA	CCTGAACTCC	TGGGGGGACC	GTCAGTCTTC	CTCTTCCCCC	1320
CAAAACCCAA	GGACACCCTC	ATGATCTCCC	GGACCCCTGA	GGTCACATGC	GTGGTGGTGG	1380
ACGTGAGCCA	CGAAGACCCT	GAGGTCAAGT	TCAACTGGTA	CGTGGACGGC	GTGGAGGTGC	1440
ATAATGCCAA	GACAAAGCCG	CGGGAGGAGC	AGTACAACAG	CACGTACCGT	GTGGTCAGCG	1500
TCCTCACCGT	CCTGCACCAG	GA CTGGCTGA	ATGGCAAGGA	GTACAAGTGC	AAGGTCTCCA	1560
ACAAAGCCCT	CCCAGCCCCC	ATCGAGAAAA	CCATCTCCAA	AGCCAAAGGT	GGGACCCGTG	1620
GGGTGCGAGG	GCCACATGGA	CAGAGGCCGG	CTCGGCCAC	CCTCTGCCCT	GAGAGTGACC	1680
GCTGTACCAA	CCTCTGTCCC	TACAGGGCAG	CCCCGAGAAC	CACAGGTGTA	CACCCTGCCC	1740
CCATCCCGGG	ATGAGCTGAC	CAAGAACCAG	GTCAGCCTGA	CCTGCCTGGT	CAAAGGCTTC	1800
TATCCCAGCG	ACATCGCCGT	GGAGTGGGAG	AGCAATGGGC	AGCCGGAGAA	CAACTACAAG	1860
ACCACGCCTC	CCGTGCTGGA	CTCCGACGGC	TCCTTCTTCC	TCTACAGCAA	GCTCACCGTG	1920
GACAAGAGCA	GGTGGCAGCA	GGGGAACGTC	TTCTCATGCT	CCGTGATGCA	TGAGGCTCTG	1980
CACAACCACT	ACACGCAGAA	GAGCCTCTCC	CTGTCTCCGG	GTAAATGAGT	GCGACGGCCG	2040
GCAAGCCCCC	GCTCCCCGGG	CTCTCGCGGT	CGCACGAGGA	TGCTTGGCAC	GTACCCCCTG	2100
TACATACTTC	CCGGGCGCCC	AGCATGGAAA	TAAAGCACCC	AGCGCTGCCC	TGGGCCCCTG	2160
CGAGACTGTG	ATGGTTCTTT	CCACGGGTCA	GGCCGAGTCT	GAGGCCTGAG	TGGCATGAGG	2220
GAGGCAGAGC	GGGTCCCACT	GTCCCCACAC	TGGCCCAGGC	TGTGCAGGTG	TGCCTGGGCC	2280
CCCTAGGGTG	GGGCTCAGCC	AGGGGCTGCC	CTCGGCAGGG	TGGGGGATTT	GCCAGCGTGG	2340
CCCTCCCTCC	AGCAGCACCT	GCCCTGGGCT	GGGCCACGGG	AAGCCCTAGG	AGCCCCTGGG	2400
GACAGACACA	CAGCCCCTGC	CTCTGTAGGA	GACTGTCCTG	TTCTGTGAGC	GCCCCTGTCC	2460
TCCCGACCTC	CATGCCCACT	CGGGGGCATG	CCTGCAGGTC	GACTCTAGAG	GATCCCCGGG	2520
TACCGAGCTC	GAATTCATCG	ATGATATCAG	ATCTGCCGGT	CTCCCTATAG	TGAGTCGTAT	2580
TAATTTTCGAT	AAGCCAGGTT	AACCTGCATT	AATGAATCGG	CCAACGCGCG	GGGAGAGGCG	2640
GTTTGCGTAT	TGGGCGCTCT	TCCGCTTCCT	CGCTCACTGA	CTCGCTGCGC	TCGGTCGTTC	2700

GGCTGCGGCG	AGCGGTATCA	GCTCACTCAA	AGGCGGTAAT	ACGGTTATCC	ACAGAATCAG	2760
GGGATAACGC	AGGAAAGAAC	ATGTGAGCAA	AAGGCCAGCA	AAAGGCCAGG	AACCGTAAAA	2820
AGGCCGCGTT	GCTGGCGTTT	TTCCATAGGC	TCCGCCCCCC	TGACGAGCAT	CACAAAAATC	2880
GACGCTCAAG	TCAGAGGTGG	CGAAACCCGA	CAGGACTATA	AAGATACCAG	GCGTTTCCCC	2940
CTGGAAGCTC	CCTCGTGCGC	TCTCCTGTTC	CGACCCTGCC	GCTTACCGGA	TACCTGTCCG	3000
CCTTTCTCCC	TTCGGGAAGC	GTGGCGCTTT	CTCAATGCTC	ACGCTGTAGG	TATCTCAGTT	3060
CGGTGTAGGT	CGTTCGCTCC	AAGCTGGGCT	GTGTGCACGA	ACCCCCGTT	CAGCCCGACC	3120
GCTGCGCCTT	ATCCGGTAAC	TATCGTCTTG	AGTCCAACCC	GGTAAGACAC	GACTTATCGC	3180
CACTGGCAGC	AGCCACTGGT	AACAGGATTA	GCAGAGCGAG	GTATGTAGGC	GGTGCTACAG	3240
AGTTCTTGAA	GTGGTGGCCT	AACTACGGCT	ACACTAGAAG	GACAGTATTT	GGTATCTGCG	3300
CTCTGCTGAA	GCCAGTTACC	TTCGGAAAAA	GAGTTGGTAG	CTCTTGATCC	GGCAAACAAA	3360
CCACCGCTGG	TAGCGGTGGT	TTTTTTGTTT	GCAAGCAGCA	GATTACGCGC	AGAAAAAAG	3420
GATCTCAAGA	AGATCCTTTG	ATCTTTTCTA	CGGGGTCTGA	CGCTCAGTGG	AACGAAAAC	3480
CACGTTAAGG	GATTTTGGTC	ATGAGATTAT	CAAAAAGGAT	CTTCACCTAG	ATCCTTTTAA	3540
ATTAAAAATG	AAGTTTTTAA	TCAATCTAAA	GTATATATGA	GTAAACTTGG	TCTGACAGTT	3600
ACCAATGCTT	AATCAGTGAG	GCACCTATCT	CAGCGATCTG	TCTATTTCTG	TCATCCATAG	3660
TTGCCTGACT	CCCCGTCTGT	TAGATAACTA	CGATACGGGA	GGGCTTACCA	TCTGGCCCCA	3720
GTGCTGCAAT	GATACGCGA	GACCCACGCT	CACCGGCTCC	AGATTTATCA	GCAATAAACC	3780
AGCCAGCCGG	AAGGGCCGAG	CGCAGAAAGT	GTCCTGCAAC	TTTATCCGCC	TCCATCCAGT	3840
CTATTAATTG	TTGCCGGGAA	GCTAGAGTAA	GTAGTTCGCC	AGTTAATAGT	TTGCGCAACG	3900
TTGTTGCCAT	TGCTACAGGC	ATCGTGGTGT	CACGCTCGTC	GTTTGGTATG	GCTTCATTCA	3960
GCTCCGGTTC	CCAACGATCA	AGGCGAGTTA	CATGATCCCC	CATGTTGTGC	AAAAAAGCGG	4020
TTAGCTCCTT	CGGTCCTCCG	ATCGTTGTCA	GAAGTAAGTT	GGCCGCAGTG	TTATCACTCA	4080
TGGTTATGGC	AGCACTGCAT	AATTCTCTTA	CTGTCATGCC	ATCCGTAAGA	TGCTTTTCTG	4140
TGACTGGTGA	GTACTCAACC	AAGTCATTCT	GAGAATAGTG	TATGCGGCGA	CCGAGTTGCT	4200
CTTGCCCGGC	GTCAATACGG	GATAATACCG	CGCCACATAG	CAGAACTTTA	AAAGTGCTCA	4260
TCATTGGAAG	ACGTTCTTCG	GGGCGAAAAA	TCTCAAGGAT	CTTACCGCTG	TTGAGATCCA	4320
GTTTCGATGA	ACCCACTCGT	GCACCCAACT	GATCTTCAGC	ATCTTTTACT	TTCACCAGCG	4380
TTTCTGGGTG	AGCAAAAACA	GGAAGGCAAA	ATGCCGCAAA	AAAGGGAATA	AGGGCGACAC	4440
GGAAATGTTG	AATACTCATA	CTCTTCCTTT	TTCAATATTA	TTGAAGCATT	TATCAGGGTT	4500
ATTGTCTCAT	GAGCGGATAC	ATATTTGAAT	GTATTTAGAA	AAATAAACAA	ATAGGGGTTC	4560

CGCGCACATT TCCCCGAAAA GTGCCACCTG ACGTCTAAGA AACCATTATT ATCATGACAT	4620
TAACCTATAA AAATAGGCGT ATCACGAGGC CCTTTCGTCT CGCGCGTTTC GGTGATGACG	4680
GTGAAAACCT CTGACACATG CAGCTCCCGG AGACGGTCAC AGCTTGTCTG TAAGCGGATG	4740
CCGGGAGCAG ACAAGCCCGT CAGGGCGCGT CAGCGGGTGT TGGCGGGTGT CGGGGCTGGC	4800
TTAACTATGC GGCATCAGAG CAGATTGTAC TGAGAGTGCA CCATATGGAC ATATTGTCGT	4860
TAGAACGCGG CTACAATTAA TACATAACCT TATGTATCAT ACACATACGA TTTAGGTGAC	4920
ACTATA	4926

(2) INFORMATION FOR SEQ ID NO:269:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 246 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGA	246

(2) INFORMATION FOR SEQ ID NO:270:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:

TACTGGTACT TCGATCTCTG GGGCCGTGGC ACCCTGGTCA CTGTCTCCTC AG	52
---	----

(2) INFORMATION FOR SEQ ID NO:271:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 307 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

TCTCTGAAGA TCTCCTGTAA GGGCTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCGAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTCGAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACAGG GGGGGGATAG GTACTTCGAT CTCTGGGGCC GTGGCACCT GGTCACTGTC	300
TCCTCAG	307

(2) INFORMATION FOR SEQ ID NO:272:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 46 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

GCTTTTGATA TCTGGGGCCA AGGGACAATG GTCACCGTCT CTCAG	46
---	----

(2) INFORMATION FOR SEQ ID NO:273:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 310 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:273:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACATT GGCTAAATGG GGATGCTTTT GATATCTGGG GCCAAGGGAC AATGGTCACC	300
GTCTCTTCAG	310

(2) INFORMATION FOR SEQ ID NO:274:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 325 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGA	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AGGGCCTCGG ACAGTGTCAT GTATTACTGT	240
GCGAGACGGG ATTACGATAT TTTGACTGGT TATTATGCGG CTTTGTATAT CTGGGGCCAA	300
GGGACAATGG TCACCGTCTC TTCAG	325

(2) INFORMATION FOR SEQ ID NO:275:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 46 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

TACTTTGACT ACTGGGGCCA GGGAACCCTG GTCACCGTCT CCTCAG	46
--	----

(2) INFORMATION FOR SEQ ID NO:276:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 313 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTC CCATCTACTG GATCGGCTGG	60
---	----

GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCTCCA TCTCAGTCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGAGTGG TTCGGGGATT TATTATTTAC TTTGACTACT GGGGCCAGGG AACCTGGTC	300
ACCGTCTCCT CAG	313

(2) INFORMATION FOR SEQ ID NO:277:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 330 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGTTTTT CCGACTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GACACCAGAT ACAGCCCGTC CTTCCAGGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AACACCGCCT TCCTGCAGTG GAACACCCTG GAGGCTTCGG ACACCGCCAT GTATTACTGT	240
GCGAGAGGGT ATTATTATGA TTCGGGGACT TATTATAAGT CTACCCCTTT GACTATTGGG	300
GCCAGGGAAC CCTGGTCACC GTCTCCTCAG	330

(2) INFORMATION FOR SEQ ID NO:278:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 301 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACTAA CTGGCCTCTT TAACTACTGG GGCCAGGGAA CCCTGGTCAC CGTCTCCTCA	300

(2) INFORMATION FOR SEQ ID NO:279:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 298 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:279:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACATC TTTACTTTGA CTTACTGGGGC CAGGGAACCC AGGTCACCGT CTCCTCAG	298

(2) INFORMATION FOR SEQ ID NO:280:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 298 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGTTTTA GCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACATC TTTACTTTGA CTTACTGGGGC CAGGGAACCC AGGTCACCGT CTCCTCAG	298

(2) INFORMATION FOR SEQ ID NO:281:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 298 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:281:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGTACCGCCT ACCTGCAGTG GAGCAGCCTG AGGGCCTCGG ACACCGCCAT TTATTACTGT	240
GCGAGACATC TTTACTTTGA CTACTGGGGC CAGGGAACCC AGGTCACCGT CTCCTCAG	298

(2) INFORMATION FOR SEQ ID NO:282:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 298 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:282:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATCGCTGT	240
GCGAGACATC TTTACTTTGA CTACTGGGGC CAGGGAACCC AGGTCACCGT CTCCTCAG	298

(2) INFORMATION FOR SEQ ID NO:283:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 295 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:283:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCCAATG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTGGCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180

AGTACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACAAG GGTTTGACTA CTGGGGCCAG GGAACCCTGG TCACCGTCTC CTCAG 295

(2) INFORMATION FOR SEQ ID NO:284:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 295 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:284:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAACTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACAAA CTTTGTGACTA CTGGGGCCAG GGAACCCTGG TCACCGTCTC CTCAG 295

(2) INFORMATION FOR SEQ ID NO:285:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 316 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:285:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGGACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACATG GTATAGCAGC AGCTGGTACG TGGTTCGACC CCTGGGGCCA GGAACCCTG 300
 GTCACCGTCT CCTCAG 316

(2) INFORMATION FOR SEQ ID NO:286:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 310 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:286:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGGCCGGGT ATACCAGCAG CTGGTTCTTT GACTTCTGGG GCCAGGGAAC CCTGGTCACC	300
GTCTCCTCAG	310

(2) INFORMATION FOR SEQ ID NO:287:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 298 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:287:

TCTCTGAAGA TCTCCTGTAG GGGTTCTGGA TACAGCTTTT CCAGTTACTG GATCGCCTGG	60
GTGCGCCAGA TGCCCGGGA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GAAACCAGAT ACAGTCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACAGG GCTACTTTGA CTA CTGGGGC CAGGGAACCC TGGTCACCGT CTCCTCAG	298

(2) INFORMATION FOR SEQ ID NO:288:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 300 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:288:

TCTCTGAAGA TCTCCTGTAA GGTTTCTGGA TACAGCTTAA CCAGTTATTG GATCGGCTGG	60
---	----

GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACAAA GGGGTACTTT GACTACTGGG GCCAGGGAAC CCTGGTCACC GTCTCCTCAG 300
 (2) INFORMATION FOR SEQ ID NO:289:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 301 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:289:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGGGGAT CGTGGTACTT TGACTACTGG GGCCAGGGAA CCCTGGTCAC CGTCTCCTCA 300
 G 301

(2) INFORMATION FOR SEQ ID NO:290:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 49 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:290:

AACTGGTTCG ACCCCTGGGG CCAGGGAACC CTGGTCACCG TCTCCTCAG 49

(2) INFORMATION FOR SEQ ID NO:291:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 307 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:291:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAACCTTA CCACCTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCGTC 180
 AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACTCC CCAATGACAG TTGGTTCGAC CCCTGGGGCC AGGAACCCT GGTCACCGTC 300
 TCCTCAG 307

(2) INFORMATION FOR SEQ ID NO:292:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 325 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:292:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACGGG GGTACTATGG TTCGGGGAGT TATTATAACT GGTTCGACCC CTGGGGCCAG 300
 GGAACCCTGG TCACCGTCTC CTCAG 325

(2) INFORMATION FOR SEQ ID NO:293:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 61 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:293:

TACTACTACT ACTACGGTAT GGACGTCTGG GGGCAAGGGA CCACGGTCAC CGTCTCCTCA 60
 G 61

(2) INFORMATION FOR SEQ ID NO:294:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 325 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:294:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAACTTTA TCACCTACTG GATCGGCTGG	60
GTGCGCCAGA TACCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACATG AGCAGCTGGT ACAGGGTTAC TACTACTACG GTATGGACGT CTGGGGGCAA	300
GGGACCACGG TCACCGTCTC CTCAG	325

(2) INFORMATION FOR SEQ ID NO:295:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 322 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:295:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACACCTTTA CCAGTTACTG GATCGCCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGAGATA TGGGGGGGGC CTCCTACTTC TACTTCGGTA TGGACGTCTG GGGGCAAGGG	300
ACCACGGTCA CCGTCTCCTC AG	322

(2) INFORMATION FOR SEQ ID NO:296:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 307 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:296:

```

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTG CCAACTACTG GATCGGCTGG      60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTTTCC TGGTGACTCT      120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAACTCCATC      180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT      240
GCGAGACACC ACGACTACTA CGGTATGGAC GTCTGGGGGC AAGGGACCAC GGTCACCGTC      300
TCCTCAG                                           307

```

(2) INFORMATION FOR SEQ ID NO:297:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 301 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:297:

```

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG      60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTTTCC TGGTGACTCT      120
GATACCAGAT ACAGCCCGCC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC      180
AACACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT      240
GCGAGACGCT ACTACGGTAT GGACGTCTGG GGGCAAGGGA CCACGGTCAC CGTCTCCTCA      300
G                                           301

```

(2) INFORMATION FOR SEQ ID NO:298:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 812 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(199..247, 419..714)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:298:

```

TTTTCTGGCC TGACAACCAG GGTGGCGCAG GATGCTCAGT GCAGAGAGGA AGAAGCAGGT      60

```

GGTCTCTGCA GCTGGAAGCT CAGCTCCCAC CCAGCTGCTT TGCATGTCCC TCCCAGCTGC 120

CCTACCTTCC AGAGCCCATATA TCAATGCCTG TGTCAGAGCC CTGGGGAGGA ACTGCTCAGT 180

TAGGACCCAG AGGGAACC ATG GAA GCC CCA GCT CAG CTT CTC TTC CTC CTG 231
Met Glu Ala Pro Ala Gln Leu Leu Phe Leu Leu
1 5 10

CTA CTC TGG CTC CCA G GTGAGGGGGA ACCATGAGGT GGTTTTGCAC 277
Leu Leu Trp Leu Pro
15

ATTAGTGAAA ACTCTTGCCA CCTCTGCTCA GCAAGAAATA TAATTAAAAT TCAAAGTATA 337

TCAACAATTT TGGCTCTACT CAAAGACAGT TGTTTTGATC TTGATTACAT GAGTGCATTT 397

CTGTTTTATT TCCAATTTCA G AT ACC ACC GGA GAA ATT GTG TTG ACA CAG 447
Asp Thr Thr Gly Glu Ile Val Leu Thr Gln
20 25

TCT CCA GCC ACC CTG TCT TTG TCT CCA GGG GAA AGA GCC ACC CTC TCC 495
Ser Pro Ala Thr Leu Ser Leu Ser Pro Gly Glu Arg Ala Thr Leu Ser
30 35 40

TGC AGG GCC AGT CAG AGT GTT AGC AGC TAC TTA GCC TGG TAC CAA CAG 543
Cys Arg Ala Ser Gln Ser Val Ser Ser Tyr Leu Ala Trp Tyr Gln Gln
45 50 55

AAA CCT GGC CAG GCT CCC AGG CTC CTC ATC TAT GAT GCA TCC AAC AGG 591
Lys Pro Gly Gln Ala Pro Arg Leu Leu Ile Tyr Asp Ala Ser Asn Arg
60 65 70

GCC ACT GGC ATC CCA GCC AGG TTC AGT GGC AGT GGG TCT GGG ACA GAC 639
Ala Thr Gly Ile Pro Ala Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp
75 80 85 90

TTC ACT CTC ACC ATC AGC AGC CTA GAG CCT GAA GAT TTT GCA GTT TAT 687
Phe Thr Leu Thr Ile Ser Ser Leu Glu Pro Glu Asp Phe Ala Val Tyr
95 100 105

TAC TGT CAG CAG CGT AGC AAC TGG CCT CCCACAGTGA TTCCACATGA 734
Tyr Cys Gln Gln Arg Ser Asn Trp Pro
110 115

AACAAAAACC CCAACAAGAC CATCAGTGTT TACTAGATTA TTATACCAGC TGCTTCCTTT 794

ACAGACAGCT AGTGGGGT 812

(2) INFORMATION FOR SEQ ID NO:299:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 115 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:299:

Met Glu Ala Pro Ala Gln Leu Leu Phe Leu Leu Leu Leu Trp Leu Pro
 1 5 10 15
 Asp Thr Thr Gly Glu Ile Val Leu Thr Gln Ser Pro Ala Thr Leu Ser
 20 25 30
 Leu Ser Pro Gly Glu Arg Ala Thr Leu Ser Cys Arg Ala Ser Gln Ser
 35 40 45
 Val Ser Ser Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro
 50 55 60
 Arg Leu Leu Ile Tyr Asp Ala Ser Asn Arg Ala Thr Gly Ile Pro Ala
 65 70 75 80
 Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser
 85 90 95
 Ser Leu Glu Pro Glu Asp Phe Ala Val Tyr Tyr Cys Gln Gln Arg Ser
 100 105 110
 Asn Trp Pro
 115

(2) INFORMATION FOR SEQ ID NO:300:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 900 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(180..228, 398..693)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:300:

AGGGCGGCGC AGATGCTCAG TGCAGAGAGA AGAAACAGGT GGTCTCTGCA GCTGGAAGCT 60
 CAGCTCCAC CCCAGCTGCT TTGCATGTCC CTCCCAGCTG CCCTACCTTC CAGAGCCCAT 120
 ATCAATGCCT GGGTCAGAGC TCTGGGGAGG AACTGCTCAG TTAGGACCCA GACGGAACC 179
 ATG GAA GCC CCA GCG CAG CTT CTC TTC CTC CTG CTA CTC TGG CTC ACA G 228
 Met Glu Ala Pro Ala Gln Leu Leu Phe Leu Leu Leu Trp Leu Thr
 1 5 10 15
 GTGAGGGGAA TATGAGGTGT CTTTGCACAT CAGTGAAAAC TCCTGCCACC TCTGCTCAGC 288
 AAGAAATATA ATTAAAATTC AAAATAGATC AACAATTTTG GCTCTACTCA AAGACAGTGG 348
 GTTTGATTTT GATTACATGA GTGCATTTCT GTTTTATTTT CAATTTTCTG AT ACC 402

Asp Thr

ACC GGA GAA ATT GTG TTG ACA CAG TCT CCA GCC ACC CTG TCT TTG TCT	450
Thr Gly Glu Ile Val Leu Thr Gln Ser Pro Ala Thr Leu Ser Leu Ser	
20 25 30	
CCA GGG GAA AGA GCC ACC CTC TCC TGC AGG GCC AGT CAG GGT GTT AGC	498
Pro Gly Glu Arg Ala Thr Leu Ser Cys Arg Ala Ser Gln Gly Val Ser	
35 40 45 50	
AGC TAC TTA GCC TGG TAC CAG CAG AAA CCT GGC CAG GCT CCC AGG CTC	546
Ser Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro Arg Leu	
55 60 65	
CTC ATC TAT GAT GCA TCC AAC AGG GCC ACT GGC ATC CCA GCC AGG TTC	594
Leu Ile Tyr Asp Ala Ser Asn Arg Ala Thr Gly Ile Pro Ala Arg Phe	
70 75 80	
AGT GGC AGT GGG CCT GGG ACA GAC TTC ACT CTC ACC ATC AGC AGC CTA	642
Ser Gly Ser Gly Pro Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu	
85 90 95	
GAG CCT GAA GAT TTT GCA GTT TAT TAC TGT CAG CAG CGT AGC AAC TGG	690
Glu Pro Glu Asp Phe Ala Val Tyr Tyr Cys Gln Gln Arg Ser Asn Trp	
100 105 110	
CAT CCCACAGTGA TTCCACATGA AACAAAAACC CCAACAAGAC CATCAGTGTT	743
His	
115	
TACTAGATTA TTATACCAGC TGCTTCCTTT ACAGACAGCT AGTGGGGTGG CCACTCAGTG	803
TTAGCATCTC AGCTCTATTT GGCCATTTTG GAGTTCAAGT TGTCAAGTCC AAAATTACTT	863
ATGTTAGTCC ATTGCATCAT ACCATTTTCAG TGTGGCT	900

(2) INFORMATION FOR SEQ ID NO:301:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 115 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:301:

Met Glu Ala Pro Ala Gln Leu Leu Phe Leu Leu Leu Leu Trp Leu Thr	
1 5 10 15	
Asp Thr Thr Gly Glu Ile Val Leu Thr Gln Ser Pro Ala Thr Leu Ser	
20 25 30	
Leu Ser Pro Gly Glu Arg Ala Thr Leu Ser Cys Arg Ala Ser Gln Gly	
35 40 45	
Val Ser Ser Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro	

(2) INFORMATION FOR SEQ ID NO:302:

(A) LENGTH: 900 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: join(116..164, 352..650)

CCGCCCCAGC	TGCTTTGCAT	GTCCCTCCCA	GCCGCCCTGC	AGTCCAGAGC	CCATATCAAT		60
GCCTGGGTCA	GAGCTCTGGA	GAAGAGCTGC	TCAGTTAGGA	ACCCCAGAGG	GAACC	ATG	118
						Met	
						1	
GAA ACC CCA GCG CAG CTT CTC TTC CTC CTG CTA CTC TGG CTC CCA G							164
Glu Thr Pro Ala Gln Leu Leu Phe Leu Leu Leu Leu Trp Leu Pro							
	5					10	
						15	
GTGAGGGGAA	CATGGGATGG	TTTTGCATGT	CAGTGAAAAC	CCTCTCAAGT	CCTGTTACCT		224
GGCAACTCTG	CTCAGTCAAT	ACAATAATTA	AAGCTCAATA	TAAAGCAATA	ATTCTGGCTC		284
TTCTGGGAAG	ACAATGGGTT	TGATTTAGAT	TACATGGGTG	ACTTTTCTGT	TTTATTTCCA		344
ATCTCAG	AT ACC ACC GGA GAA ATT GTG TTG ACG CAG TCT CCA GGC ACC						392
	Asp Thr Thr Gly Glu Ile Val Leu Thr Gln Ser Pro Gly Thr						
		20				25	
							30
CTG TCT TTG TCT CCA GGG GAA AGA GCC ACC CTC TCC TGC AGG GCC AGT							440
Leu Ser Leu Ser Pro Gly Glu Arg Ala Thr Leu Ser Cys Arg Ala Ser							
	35					40	
							45
CAG AGT GTT AGC AGC AGC TAC TTA GCC TGG TAC CAG CAG AAA CCT GGC							488
Gln Ser Val Ser Ser Ser Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly							
	50					55	
							60
CAG GCT CCC AGG CTC CTC ATC TAT GGT GCA TCC AGC AGG GCC ACT GGC							536

Gln	Ala	Pro	Arg	Leu	Leu	Ile	Tyr	Gly	Ala	Ser	Ser	Arg	Ala	Thr	Gly		
	65						70					75					
ATC	CCA	GAC	AGG	TTC	AGT	GGC	AGT	GGG	TCT	GGG	ACA	GAC	TTC	ACT	CTC	584	
Ile	Pro	Asp	Arg	Phe	Ser	Gly	Ser	Gly	Ser	Gly	Thr	Asp	Phe	Thr	Leu		
	80					85					90						
ACC	ATC	AGC	AGA	CTG	GAG	CCT	GAA	GAT	TTT	GCA	GTG	TAT	TAC	TGT	CAG	632	
Thr	Ile	Ser	Arg	Leu	Glu	Pro	Glu	Asp	Phe	Ala	Val	Tyr	Tyr	Cys	Gln		
	95				100					105					110		
CAG	TAT	GGT	AGC	TCA	CCT	CCCACAGTGA	TTCAGCTTGA	AACAAAAACC								680	
Gln	Tyr	Gly	Ser	Ser	Pro												
				115													
TCTGCAAGAC	CTTCATTGTT	TACTAGATTA	TACCAGCTGC	TTCCTTTACA	GATAGCTGCT											740	
GCAATGACAA	CTCAATTTAG	CATCTCTCTC	TGCTTGGGCA	TTTTGGGGAT	CTTAAAAAAG											800	
TAATCCCTTG	ATATATTTTT	GACTCTGATT	CCTGCATTTT	TCCTCAGACC	AAGATGGACA											860	
GCCAGGTTTA	AGCACAGTTT	CACAGTAATG	GCCACTGGAT													900	

(2) INFORMATION FOR SEQ ID NO:303:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 116 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:303:

Met	Glu	Thr	Pro	Ala	Gln	Leu	Leu	Phe	Leu	Leu	Leu	Leu	Trp	Leu	Pro		
1				5				10					15				
Asp	Thr	Thr	Gly	Glu	Ile	Val	Leu	Thr	Gln	Ser	Pro	Gly	Thr	Leu	Ser		
			20					25					30				
Leu	Ser	Pro	Gly	Glu	Arg	Ala	Thr	Leu	Ser	Cys	Arg	Ala	Ser	Gln	Ser		
		35					40					45					
Val	Ser	Ser	Ser	Tyr	Leu	Ala	Trp	Tyr	Gln	Gln	Lys	Pro	Gly	Gln	Ala		
	50					55					60						
Pro	Arg	Leu	Leu	Ile	Tyr	Gly	Ala	Ser	Ser	Arg	Ala	Thr	Gly	Ile	Pro		
	65				70					75					80		
Asp	Arg	Phe	Ser	Gly	Ser	Gly	Ser	Gly	Thr	Asp	Phe	Thr	Leu	Thr	Ile		
				85					90					95			
Ser	Arg	Leu	Glu	Pro	Glu	Asp	Phe	Ala	Val	Tyr	Tyr	Cys	Gln	Gln	Tyr		
		100						105					110				
Gly	Ser	Ser	Pro														
			115														

(2) INFORMATION FOR SEQ ID NO:304:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 847 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(226..280, 406..701)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:304:

```

AAACACATTC TCTGCAGACA AATTTGAGCT ACCTTGATCT TACCTGGACA GGTGGGGACA      60
CTGAGCTGGT GCTGAGTTAC TCAGATGCGC CAGCTCTGCA GCTGTGCCCC GCCTGCCCCA      120
TCCCCTGCTC ATTTGCATGT TCCCAGAGCA CAACCTCCTG CCCTGAAGCC TTATTAATAG      180
GCTGGTCAGA CTTTGTGCAG GAATCAGACC CAGTCAGGAC ACAGC ATG GAC ATG      234
                                   Met Asp Met
                                   1
AGG GTC CTC GCT CAG CTC CTG GGG CTC CTG CTG CTC TGT TTC CCA G      280
Arg Val Leu Ala Gln Leu Leu Gly Leu Leu Leu Leu Cys Phe Pro
      5              10              15
GTAAGGATGG AGAACACTAG CAGTTTACTC AGCCCAGGGT GCTCAGTACT GCTTTACTAT      340
TCAGGGAAAT TCTCTTACAA CATGATTAAT TGTGTGGACA TTTGTTTTTA TGTTTCCAAT      400
CTCAG GT GCC AGA TGT GAC ATC CAG ATG ACC CAG TCT CCA TCC TCA      446
      Gly Ala Arg Cys Asp Ile Gln Met Thr Gln Ser Pro Ser Ser
      20              25              30
CTG TCT GCA TCT GTA GGA GAC AGA GTC ACC ATC ACT TGT CGG GCG AGT      494
Leu Ser Ala Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ala Ser
      35              40              45
CAG GGT ATT AGC AGC TGG TTA GCC TGG TAT CAG CAG AAA CCA GAG AAA      542
Gln Gly Ile Ser Ser Trp Leu Ala Trp Tyr Gln Gln Lys Pro Glu Lys
      50              55              60
GCC CCT AAG TCC CTG ATC TAT GCT GCA TCC AGT TTG CAA AGT GGG GTC      590
Ala Pro Lys Ser Leu Ile Tyr Ala Ala Ser Ser Leu Gln Ser Gly Val
      65              70              75              80
CCA TCA AGG TTC AGC GGC AGT GGA TCT GGG ACA GAT TTC ACT CTC ACC      638
Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr
      85              90              95
ATC AGC AGC CTG CAG CCT GAA GAT TTT GCA ACT TAT TAC TGC CAA CAG      686
Ile Ser Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln
      100              105              110
TAT AAT AGT TAC CCA CCCACAGTGT TACACACCCA AACATAAACC CCCAGGGAAG      741

```


Tyr Asn Ser Tyr Pro
115

CAGATGTGTG AGGCTGGGCT GCCCCAGCTG CTTCTCCTGA TGCCTCCATC AGCTGAGAGT 801

GTTCTCCTCAGA TGCAGCCACA CTCTGATGGT GTTGGTAGAT GGGGAC 847

(2) INFORMATION FOR SEQ ID NO:305:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 117 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:305:

Met	Asp	Met	Arg	Val	Leu	Ala	Gln	Leu	Leu	Gly	Leu	Leu	Leu	Cys		
1				5				10					15			
Phe	Pro	Gly	Ala	Arg	Cys	Asp	Ile	Gln	Met	Thr	Gln	Ser	Pro	Ser	Ser	
			20					25					30			
Leu	Ser	Ala	Ser	Val	Gly	Asp	Arg	Val	Thr	Ile	Thr	Cys	Arg	Ala	Ser	
			35				40					45				
Gln	Gly	Ile	Ser	Ser	Trp	Leu	Ala	Trp	Tyr	Gln	Gln	Lys	Pro	Glu	Lys	
			50			55					60					
Ala	Pro	Lys	Ser	Leu	Ile	Tyr	Ala	Ala	Ser	Ser	Leu	Gln	Ser	Gly	Val	
			65		70					75					80	
Pro	Ser	Arg	Phe	Ser	Gly	Ser	Gly	Ser	Gly	Thr	Asp	Phe	Thr	Leu	Thr	
				85					90					95		
Ile	Ser	Ser	Leu	Gln	Pro	Glu	Asp	Phe	Ala	Thr	Tyr	Tyr	Cys	Gln	Gln	
			100					105					110			
Tyr	Asn	Ser	Tyr	Pro												
			115													

(2) INFORMATION FOR SEQ ID NO:306:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 134 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:306:

GCCTCGGACA CCGCCATGTA TTACTGTGTG AGACATTTAT GGTTCGGGGA GTTACGCGGT 60

GTGAACGTCT GGGGCCAAGG GACCACGGTC ACCGTCTCCT CAGCCAAAAC GACACCCCCA 120

TCTGTCTATC CACT

134

(2) INFORMATION FOR SEQ ID NO:307:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 125 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:307:

GCCTCGGACA CCGCCATGTA TTACTGTGCG AGACACTGGG CATTGGATGC TCTTGATGTC 60
 TGGGGCCAAG GGACAATGCT CACCGTCTCT TCAGCCAAAA CGACACCCCC ATCTGTCTAT 120
 CCACT 125

(2) INFORMATION FOR SEQ ID NO:308:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 122 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:308:

GCCTCGGACA CCGCCATGTA TTACTGTGCG AGAACTGGGG ATGATGCTTT TGATATCTGG 60
 GGCCAAGGGA CAATGGTCAC CGTCTCTTCA GCCAAAACAA CACCCCCATC AGTCTATCCA 120
 CT 122

(2) INFORMATION FOR SEQ ID NO:309:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 122 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:309:

GACTCGGACA CCGCCATGTA TTACTGTGCG AGACAGGGGA GAGATGCTTT AGATATCTGG 60
 GGCCAAGGGA CAATGGTCAC CGTCTCTTCA GCCAAAACAA CACCCCCATC AGTCTATCCA 120
 CT 122

(2) INFORMATION FOR SEQ ID NO:310:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 137 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:310:

GCCTCGGACA CCGCCATGTA TTATTGTGTG AGACATAGGG ACTATATTTC GGGGAGTTAT	60
TTTCCTGACT ACTGGGGCCA GGGAACCCTG GTCACCGTCT CCTCAGCCAA AACAAACACCC	120
CCATCAGTCT ATCCACT	137

(2) INFORMATION FOR SEQ ID NO:311:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 122 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:311:

GCCTCGGACA CCGCCATGTA TTA CTGTGCG AGAACTGGGG ATGATGCTTT TGATATCTGG	60
GGCCAAGGGA CAATGGTCAC CGTCTCTTCA GCCAAAACAA CACCCCCATC AGTCTATCCA	120
CT	122

(2) INFORMATION FOR SEQ ID NO:312:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 116 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:312:

GCCTCGGACA CCGCCATGTA TTA CTGTGCG AGACATGGGT CTATGGATAT CTGGGGCCAA	60
GGGACAATGG TCACCGTCTC TTCAGCTACA ACAACAGCCC CATCTGTCTA TCCCTT	116

(2) INFORMATION FOR SEQ ID NO:313:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 128 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:313:

GCCTCGGACA CCGCCATGTA TTACTGTGCG AGAGAGAGCG GTCACTGGGG ATCGTTTGAC	60
TATTGGGGCC AGGGAACCCT GGTCACCGTC TCCTCAGCTA CAACAACAGC CCCATCTGTC	120
TATCCCTT	128

(2) INFORMATION FOR SEQ ID NO:314:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 125 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:314:

GCCTCGGACA CCGCCATGTA TTACTGTGCG AGAAGGGACC CCCCTGATGC TTTTGATATC	60
TGGGGCCAAG GGACAATGGT CACCGTCTCT TCAGCTACAA CAACAGCCCC ATCTGTCTAT	120
CCCTT	125

(2) INFORMATION FOR SEQ ID NO:315:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 131 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:315:

GCCTCGGACA CCGCCATGTA TTACTGTGCG AGACGGGGGC CTTACTACTA CTACGGTATG	60
GACGTCTGGG GCCAAGGGAC CACGGTCACC GTCTCCTCAG CTACAACAAC AGCCCCATCT	120
GTCTATCCCT T	131

(2) INFORMATION FOR SEQ ID NO:316:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1674 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:316:

CGAGAGGGGC GGGGGGAAGA CTACTATCCC AGGCAGGTTT TAGGTTCCAG AGTCTGCGAG	60
AAATCCCACC ATCTACCCAC TGACACTCCC ACCAGTCCTG TGCAGTGATC CCGTGATAAT	120
CGGCTGCCTG ATTCACGATT ACTTCCCTTT CGGCACGATG AATGTGACCT GGGGAAAGAG	180
TGGGAAGGAT ATAACCACCG TGAACTTTCC ACCTGCCCTC GCCTCTGGGG GACGGTACAC	240
CATGAGCAGC CAGTTAACCC TGCCAGCTGT CGAGTGCCCA GAAGGAGAGT CCGTGAAATG	300
TTCCGTGCAA CATGACTCTA ACCCCGTCCA AGAATTGGAT GTGAATTGCT CTGGTAAAGA	360
ACGTTAGGGG GTCAGCTAGG GGTGGGATAA GTCCTACCTT ATCTAGATCC ATATATCCCT	420
CTGATGCACA CCCTCACAGG AATCCCTCAG AAACCTCCAC TATGGGGATT GGGGGAAGGA	480
AGCGTAAACA GGTCTAGAAG GAGCTGGAGG CCTCAGAACA TCCAGAAACG GGGACAGCAA	540
AGGAGACAAG GAGAATATAC TGATTTGCTA GGACATCTTC TGTTACAGGT CCTACTCCTC	600
CTCCTCCTAT TACTATTCCT TCCTGCCAGC CCAGCCTGTC ACTGCAGCGG CCAGCTCTTG	660
AGGACCTGCT CCTGGGTTCA GATGCCAGCA TCACATGTAC TCTGAATGGC CTGAGAAATC	720
CTGAGGGAGC TGCTTTTACC TGGGAGCCCT CCACTGGGAA GGATGCAGTG CAGAAGAAAG	780
CTGCGCAGAA TTCCTGCGGC TGCTACAGTG TGTCCAGCGT CCTGCCTGGC TGTGCTGAGC	840
GCTGGAACAG TGGCGCATCA TTCAAGTGCA CAGTTACCCA TCCTGAGTCT GGCACCTTAA	900
CTGGCACAAT TGCCAAAGTC ACAGGTGAGC TCAGATGCAT ACCAGGACAT TGTATGACGT	960
TCCCTGCTCA CATGCCTGCT TTCTTCCTAT AATACAGATG CTCAACTAAC TGCTCATGTC	1020
CTTATATCAC AGAGGGAAAT TGGAGCTATC TGAGGAACTG CCCAGAAGGG AAGGGCAGAG	1080
GGGTCTTGCT CTCCTTGTCT GAGCCATAAC TCTTCTTTCT ACCTTCCAGT GAACACCTTC	1140
CCACCCCAGG TCCACCTGCT ACCGCCGCCG TCGGAGGAGC TGGCCCTGAA TGAGCTCTTG	1200
TCCCTGACAT GCCTGGTGCG AGCTTTCAAC CCTAAAGAAG TGCTGGTGCG ATGGCTGCAT	1260
GGAAATGAGG AGCTGTCCCC AGAAAGCTAC CTAGTGTTTG AGCCCCTAAA GGAGCCAGGC	1320
GAGGGAGCCA CCACCTACCT GGTGACAAGC GTGTTGCGTG TATCAGCTGA AACCTGGAAA	1380
CAGGGTGACC AGTACTCCTG CATGGTGGGC CACGAGGCCT TGCCCATGAA CTTACCCAG	1440

AAGACCATCG ACCGTCTGTC GGGTAAACCC ACCAATGTCA GCGTGTCTGT GATCATGTCA	1500
GAGGGAGATG GCATCTGCTA CTGAGCCACC CTGCCTGTCC CTACTCCTAG AATAAACTCT	1560
GTGCTCATCC AAAGTATCCC TGCACCTCCA CCCAGTGCCT GTCCACCACC CTGGGGTCTA	1620
CGAAACACAG GGAGGGGTCA GGGCCCAGGG AGGGAGAAAT ACCACCACCT AAGC	1674

(2) INFORMATION FOR SEQ ID NO:317:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 133 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:317:

GCCTCGGACA CCGCCATGTA TTACTGTGTG AGACATTTAT GGTTCGGGGA GTTACGCGGT	60
GTGAACGTCT GGGGCCAAGG GACCACGGTC ACCGTCTCCT CAGCCAAAAC GACACCCCCA	120
TCTGTCTATC CAC	133

(2) INFORMATION FOR SEQ ID NO:318:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 124 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:318:

GCCTCGGACA CCGCCATGTA TTACTGTGCG AGACACTGGG CATTGGATGC TCTTGATGTC	60
TGGGGCCAAG GGACAATGCT CACCGTCTCT TCAGCCAAAA CGACACCCCC ATCTGTCTAT	120
CCAC	124

(2) INFORMATION FOR SEQ ID NO:319:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 121 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:319:

GCCTCGGACA CCGCCATGTA TTACTGTGCG AGAACTGGGG ATGATGCTTT TGATATCTGG 60
 GGCCAAGGGA CAATGGTCAC CGTCTCTTCA GCCAAAACAA CACCCCCATC AGTCTATCCA 120
 C 121

(2) INFORMATION FOR SEQ ID NO:320:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 121 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:320:

GACTCGGACA CCGCCATGTA TTACTGTGCG AGACAGGGGA GAGATGCTTT AGATATCTGG 60
 GGCCAAGGGA CAATGGTCAC CGTCTCTTCA GCCAAAACAA CACCCCCATC AGTCTATCCA 120
 C 121

(2) INFORMATION FOR SEQ ID NO:321:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 136 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:321:

GCCTCGGACA CCGCCATGTA TTATTGTGTG AGACATAGGG ACTATATTTT GGGGAGTTAT 60
 TTTCTTGACT ACTGGGGCCA GGGAACCCTG GTCACCGTCT CCTCAGCCAA AACAACACCC 120
 CCATCAGTCT ATCCAC 136

(2) INFORMATION FOR SEQ ID NO:322:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 121 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:322:

GCCTCGGACA CCGCCATGTA TTACTGTGCG AGAACTGGGG ATGATGCTTT TGATATCTGG 60
 GGCCAAGGGA CAATGGTCAC CGTCTCTTCA GCCAAAACAA CACCCCCATC AGTCTATCCA 120
 C 121

(2) INFORMATION FOR SEQ ID NO:323:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 115 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:323:

GCCTCGGACA CCGCCATGTA TTACTGTGCG AGACATGGGT CTATGGATAT CTGGGGCCAA 60
 GGGACAATGG TCACCGTCTC TTCAGCTACA ACAACAGCCC CATCTGTCTA TCCCT 115

(2) INFORMATION FOR SEQ ID NO:324:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 127 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:324:

GCCTCGGACA CCGCCATGTA TTACTGTGCG AGAGAGAGCG GTCACTGGGG ATCGTTTGAC 60
 TATTGGGGCC AGGGAACCCT GGTCACCGTC TCCTCAGCTA CAACAACAGC CCCATCTGTC 120
 TATCCCT 127

(2) INFORMATION FOR SEQ ID NO:325:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 124 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:325:

GCCTCGGACA CCGCCATGTA TTACTGTGCG AGAAGGGACC CCCCTGATGC TTTTGATATC 60
 TGGGGCCAAG GGACAATGGT CACCGTCTCT TCAGCTACAA CAACAGCCCC ATCTGTCTAT 120
 CCCT 124

(2) INFORMATION FOR SEQ ID NO:326:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 130 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:326:

GCCTCGGACA CCGCCATGTA TTACTGTGCG AGACGGGGGC CTTACTACTA CTACGGTATG 60
 GACGTCTGGG GCCAAGGGAC CACGGTCACC GTCTCCTCAG CTACAACAAC AGCCCCATCT 120
 GTCTATCCCT 130

(2) INFORMATION FOR SEQ ID NO:327:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:327:

TACTGGTACT TCGATCTCTG GGGCCGTGGC AC 32

(2) INFORMATION FOR SEQ ID NO:328:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 293 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:328:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG 60

```

GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT      120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC      180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT      240
GCGAGACAAG GGACTGGGGA GGACTGGTAC TTCGATCTCT GGGGCCGTGG CAC              293

```

(2) INFORMATION FOR SEQ ID NO:329:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 290 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:329:

```

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG      60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT      120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC      180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT      240
GCGAGACAGA AGGGGGGAAG GGGGTACTTC GATCTCTGGG GCCGTGGCAC              290

```

(2) INFORMATION FOR SEQ ID NO:330:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 287 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:330:

```

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG      60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT      120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC      180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT      240
GCGAGAACTG GGTGGTACTG GTACTTCGAT CTCTGGGGCC GTGGCAC              287

```

(2) INFORMATION FOR SEQ ID NO:331:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:331:

GCTTTTGATA TCTGGGGCCA AGGGAC

26

(2) INFORMATION FOR SEQ ID NO:332:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 293 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:332:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GCGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACATG AACTGGGGA TCCGGGGGCT TTTGATATCT GGGGCCAAGG GAC	293

(2) INFORMATION FOR SEQ ID NO:333:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:333:

TACTTTGACT ACTGGGGCCA GGGAAC

26

(2) INFORMATION FOR SEQ ID NO:334:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 305 base pairs
 - (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:334:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGAACGG ATTACGATAT TTTGACTGGT TATTATAACC CTTTGTACTA CTGGGGCCAG	300
GGAAC	305

(2) INFORMATION FOR SEQ ID NO:335:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 284 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:335:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGTATAG CAGCAGCCCT TTTTGTACTAC TGGGGCCAGG GAAC	284

(2) INFORMATION FOR SEQ ID NO:336:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 290 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:336:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
---	----

GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGATCTC TACTATGGTT CGGGGAGTTT GACTACTGGG GCCAGGGAAC	290

(2) INFORMATION FOR SEQ ID NO:337:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 305 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:337:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACTCG ATTACGATAT TTTGACTGGT TATTATAACC CCTTTGACTA CTGGGGCCAG	300
GGAAC	305

(2) INFORMATION FOR SEQ ID NO:338:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 305 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:338:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCGGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACTCG ATTACGATAT TTTGACTGGT TATTATAACC CCTTTGACTA CTGGGGCCAG	300

GGAAC

305

(2) INFORMATION FOR SEQ ID NO:339:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 293 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:339:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGCCCGT ATAGCAGCAG CTGGTACAGG TTTGACTACT GGGGCCAGGG AAC	293

(2) INFORMATION FOR SEQ ID NO:340:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:340:

TACTACTACT ACTACGGTAT GGACGTCTGG GGCCAAGGGA C	41
---	----

(2) INFORMATION FOR SEQ ID NO:341:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 317 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:341:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180

AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGAGAGG GGGGGGTTTA TTACTATGGT TCGGGGAGTT ATTACTACTA CGGTATGGAC 300
 GTCTGGGGCC AAGGGAC 317

(2) INFORMATION FOR SEQ ID NO:342:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 305 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:342:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGAGAGG GAACTGGGGA CCATTACTGC TACTACTACG GTATGGACGT CTGGGGCCAA 300
 GGGAC 305

(2) INFORMATION FOR SEQ ID NO:343:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:343:

TACTGGTACT TCGATCTCTG GGGCCGTGGC ACCCTGGTC 39

(2) INFORMATION FOR SEQ ID NO:344:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 303 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:344:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTCG TCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACAGG GACTGGGGAT CGGGGAATGG TACTTCGATC TCTGGGGCCG TGGCACCTG	300
GTC	303

(2) INFORMATION FOR SEQ ID NO:345:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:345:

GCTTTTGATA TCTGGGGCCA AGGGACAATG GTC	33
--------------------------------------	----

(2) INFORMATION FOR SEQ ID NO:346:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 291 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:346:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACGGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGGTAAACT GGGATCGGGC TTTTGATATC TGGGGCCAAG GGACAATGGT C	291

(2) INFORMATION FOR SEQ ID NO:347:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:347:

TACTTTGACT ACTGGGGCCA GGGAACCCTG GTC 33

(2) INFORMATION FOR SEQ ID NO:348:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 297 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:348:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCGGCTACTG GATCGGCTGG	60
GTGCGCCAGA TACCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGTACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTTCTGT	240
GCGAGACATA AGGCGGGGAT CAACTACTTT GCCTACTGGG GCCAGGGAAC CCTGGTC	297

(2) INFORMATION FOR SEQ ID NO:349:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 297 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:349:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAATTCA CCGGCTACTG GATCGGCTGG	60
GTGCGCCAGA TACCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGTACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTTCTGT	240
GCGAGACATA AGGCGGGGAT CAACTACTTT GCCTACTGGG GCCAGGGAAC CCTGGTC	297

(2) INFORMATION FOR SEQ ID NO:350:

(i) SEQUENCE CHARACTERISTICS:

406

- (A) LENGTH: 297 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:350:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAACTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCG TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGAGACT GGGGAGAAGG GTATTACTTT GACTACTGGG GCCAGGGAAC CCTGGTC	297

(2) INFORMATION FOR SEQ ID NO:351:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 282 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:351:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGTTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGCA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCGCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACAAC TGTTGACTA CTGGGGCCAG GGAACCCTGG TC	282

(2) INFORMATION FOR SEQ ID NO:352:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 312 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:352:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA TCAGCTACTG GATCGGCTGG	60
---	----

GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACTGT ATTACTATGG TTCGGGGAGT TACCACAACT GGTTCGACCC CTGGGGCCAG 300
 GGAACCCTGG TC 312

(2) INFORMATION FOR SEQ ID NO:353:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 36 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:353:

AACTGGTTCG ACCCCTGGGG CCAGGGAACC CTGGTC 36

(2) INFORMATION FOR SEQ ID NO:354:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 318 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:354:

TCTCTGAAGA TCTCCTGTAA GGTTTCTGGA TACAACTTTG CCAGCTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCACATATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AACACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACAGG GGTATTACTA TGGTTCGGCA AATTATTATA ACATTGGTT CGACCCCTGG 300
 GGCCAGGGAA CCCTGGTC 318

(2) INFORMATION FOR SEQ ID NO:355:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 48 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:355:

TACTACTACT ACTACGGTAT GGACGTCTGG GGCCAAGGGA CCACGGTC

48

(2) INFORMATION FOR SEQ ID NO:356:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 297 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:356:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TATAGTTTTG CCAACTACGG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCC	120
GGTTCCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AACACCGCCT ACCTGCAATG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGCCCCCGG CGTACTACTA CTACGGTATG GACGTCTGGG GCCAAGGGAC CACGGTC	297

(2) INFORMATION FOR SEQ ID NO:357:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 156 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:357:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTC	156

(2) INFORMATION FOR SEQ ID NO:358:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 156 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:358:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGTTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTC	156

(2) INFORMATION FOR SEQ ID NO:359:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 156 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:359:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA GCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTC	156

(2) INFORMATION FOR SEQ ID NO:360:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 156 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:360:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAACTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTC	156

(2) INFORMATION FOR SEQ ID NO:361:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 156 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:361:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCACCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTC	156

(2) INFORMATION FOR SEQ ID NO:362:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 294 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:362:

TCTCTGAAGA TCTCCTGTAA GGGCTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCGAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTCGAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACAGG GGGGGGATAG GTACTTCGAT CTCTGGGGCC GTGGCACCTT GGTC	294

(2) INFORMATION FOR SEQ ID NO:363:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 297 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:363:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240

GCGAGACATT GGCTAAATGG GGATGCTTTT GATATCTGGG GCCAAGGGAC AATGGTC 297

(2) INFORMATION FOR SEQ ID NO:364:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 312 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:364:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AGGGCCTCGG ACAGTGTCAT GTATTACTGT	240
GCGAGACGGG ATTACGATAT TTTGACTGGT TATTATGCGG CTTTGTATAT CTGGGGCCAA	300
GGGACAATGG TC	312

(2) INFORMATION FOR SEQ ID NO:365:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 317 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:365:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGTTTTT CCGACTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GACACCAGAT ACAGCCCGTC CTTCCAGGGC CAGGTCTCCA TCTCAGTCGA CAAGTCCATC	180
AACACCGCCT TCCTGCAGTG GAACACCCTG GAGGCTTCGG ACACCGCCAT GTATTACTGT	240
GCGAGAGGGT ATTATTATGA TTCGGGGACT TATTATAAGT CTACCCCTTT GACTATTGGG	300
GCCAGGGAAC CCTGGTC	317

(2) INFORMATION FOR SEQ ID NO:366:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 288 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:366:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGTTTTT CCGACTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GACACCAGAT ACAGCCCGTC CTTCCAGGGC CAGGTCTCCA TCTCAGTCGA CAAGTCCATC	180
AACACCGCCT TCCTGCAGTG GAACACCCTG GAGGCTTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACTAA CTGGCCTCTT TAACTATTGG GGCCAGGGAA CCCTGGTC	288

(2) INFORMATION FOR SEQ ID NO:367:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 285 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:367:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACATC TTTACTTTGA CTACTGGGGC CAGGGAACCC AGGTC	285

(2) INFORMATION FOR SEQ ID NO:368:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:368:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCCAATG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTGGCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180

AGTACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACAAG GGT TTGACTA CTGGGGCCAG GGAACCCTGG TC 282

(2) INFORMATION FOR SEQ ID NO:369:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 285 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:369:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGTTTTA GCAGCTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACATC TTTACTTTGA CTACTGGGGC CAGGGAACCC AGGTC 285

(2) INFORMATION FOR SEQ ID NO:370:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 282 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:370:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAACTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACAAA CTTT TACTA CTGGGGCCAG GGAACCCTGG TC 282

(2) INFORMATION FOR SEQ ID NO:371:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 303 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:371:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGGACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACATG GTATAGCAGC AGCTGGTACG TGGTTCGACC CCTGGGGCCA GGGAACCCTG	300
GTC	303

(2) INFORMATION FOR SEQ ID NO:372:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 285 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:372:

TCTCTGAAGA TCTCCTGTAG GGGTTCTGGA TACAGCTTTT CCAGTTACTG GATCGCCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GAAACCAGAT ACAGTCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACAGG GCTACTTTGA CTACTGGGGC CAGGGAACCC TGGTC	285

(2) INFORMATION FOR SEQ ID NO:373:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 285 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:373:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120

GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGTACCGCCT ACCTGCAGTG GAGCAGCCTG AGGGCCTCGG ACACCGCCAT TTATTACTGT 240
 GCGAGACATC TTTACTTTGA CTACTGGGGC CAGGGAACCC AGGTC 285

(2) INFORMATION FOR SEQ ID NO:374:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 297 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:374:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGGCCGGGT ATACCAGCAG CTGGTTCTTT GACTTCTGGG GCCAGGGAAC CCTGGTC 297

(2) INFORMATION FOR SEQ ID NO:375:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 285 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:375:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATCGCTGT 240
 GCGAGACATC TTTACTTTGA CTACTGGGGC CAGGGAACCC AGGTC 285

(2) INFORMATION FOR SEQ ID NO:376:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 300 base pairs
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:376:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTC CCATCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGAGTGG TTCGGGGATT TATTATTTAC TTTGACTACT GGGGCCAGGG AACCTGGTC	300

(2) INFORMATION FOR SEQ ID NO:377

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 287 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:377:

TCTCTGAAGA TCTCCTGTAA GGTTTCTGGA TACAGCTTAA CCAGTTATTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACAAA GGGGTACTTT GACTACTGGG GCCAGGGAAC CCTGGTC	287

(2) INFORMATION FOR SEQ ID NO:378:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 288 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:378:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120

GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGCACC GCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGGGGAT CGTGGTACTT TGACTACTGG GGCCAGGGAA CCCTGGTC 288

(2) INFORMATION FOR SEQ ID NO:379:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 294 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:379:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAAC TTTA CCACCTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCGTC 180
 AGCACC GCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACTCC CCAATGACAG TTGGTTTCGAC CCCTGGGGCC AGGGAACCCT GGTC 294

(2) INFORMATION FOR SEQ ID NO:380:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 312 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:380:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG 60
 GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT 120
 GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC 180
 AGCACC GCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT 240
 GCGAGACGGG GGTACTATGG TTCGGGGAGT TATTATAACT GGTTCGACCC CTGGGGCCAG 300
 GGAACCCTGG TC 312

(2) INFORMATION FOR SEQ ID NO:381:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 312 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:381:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAACCTTA TCACCTACTG GATCGGCTGG	60
GTGCGCCAGA TACCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGACATG AGCAGCTGGT ACAGGGTTAC TACTACTACG GTATGGACGT CTGGGGCCAA	300
GGGACCACGG TC	312

(2) INFORMATION FOR SEQ ID NO:382:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 309 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:382:

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACACCTTTA CCAGTTACTG GATCGCCTGG	60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTATCC TGGTGACTCT	120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC	180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT	240
GCGAGAGATA TGGGGGGGGC CTCCTACTTC TACTTCGGTA TGGACGTCTG GGGCCAAGGG	300
ACCACGGTC	309

(2) INFORMATION FOR SEQ ID NO:383:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 294 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:383:

```

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTG CCAACTACTG GATCGGCTGG      60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTTTCC TGGTGACTCT      120
GATACCAGAT ACAGCCCGTC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAACTCCATC      180
AGCACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT      240
GCGAGACACC ACGACTACTA CGGTATGGAC GTCTGGGGCC AAGGGACCAC GGTC              294

```

(2) INFORMATION FOR SEQ ID NO:384:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 288 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:384:

```

TCTCTGAAGA TCTCCTGTAA GGGTTCTGGA TACAGCTTTA CCAGCTACTG GATCGGCTGG      60
GTGCGCCAGA TGCCCGGGAA AGGCCTGGAG TGGATGGGGA TCATCTTTCC TGGTGACTCT      120
GATACCAGAT ACAGCCCGCC CTTCCAAGGC CAGGTCACCA TCTCAGCCGA CAAGTCCATC      180
AACACCGCCT ACCTGCAGTG GAGCAGCCTG AAGGCCTCGG ACACCGCCAT GTATTACTGT      240
GCGAGACGCT ACTACGGTAT GGACGTCTGG GGCCAAGGGA CCACGGTC              288

```

(2) INFORMATION FOR SEQ ID NO:385:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 246 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:385:

```

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAGCTATGC TATGCACTGG      60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAG TGGGTGGCAG TTATATCATA TGATGGAAGC      120
AATAAATACT ACGCAGACTC CGTGAAGGGC CGATTACCA TCTCCAGAGA CAATTCCAAG      180
AACACGCTGT ATCTGCAAAT GAACAGCCTG AGAGCTGAGG ACACGGCTGT GTATTACTGT      240
GCGAGA                                             246

```

(2) INFORMATION FOR SEQ ID NO:386:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:386:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAGCTATGC TATGCACTGG	60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAC TGGGTGGCAG TTATATCATA TGATGGAACC	120
AATAAATACG ACGCAGACTC CGTGAAGGGC CGATTACCA TCTCCAGAGA CAATTCCAAG	180
AACACGCTGT ATCTGCAAAT GAACAGCCTG AGATCTGAGG ACACGGCTGT GTATTACTGT	240
GCGAGAGAGT CTTCCGGCTG GTACTIONGAT TTCTGGGGCC GTGGCACCTT GGTC	294

(2) INFORMATION FOR SEQ ID NO:387:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:387:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAGCTATGC TATGCACTGG	60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAC TGGGTGGCAG TTATATCATA TGATGGAACC	120
AATAAATACG ACGCAGACTC CGTGAAGGGC CGATTACCA TCTCCAGAGA CAATTCCAAG	180
AACACGCTGT ATCTGCAAAT GAACAGCCTG AGATCTGAGG ACACGGCTGT GTATTACTGT	240
GCGAGGAAGT CTTCCGGCTG GTACTIONGAT TTCTGGGGCC GTGGCACCTT GGTC	294

(2) INFORMATION FOR SEQ ID NO:388:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:388:


```

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA ATAAGTGTAC TATACACTGG      60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAC TGGGTGGCAG TTATATCATA TGATGGAGCC      120
AATAAATACG ACGCAGAGTC CGTGAAGGGC CGATTACCA TCTCCAGAGA CAATTCCAAG      180
AACATGCTGT ATCCGCAAAT GAACAGCCTG AGATCTGAGG ACACGGCTGT GTATTACTGT      240
GCGAGAGAGT CCTCCGGCTG GTACTTCGAT CTTTGGGGCC GTGGCACCCCT GGTC          294

```

(2) INFORMATION FOR SEQ ID NO:389:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:389:

```

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA ATAAGTGTAC TATACACTGG      60
GTCCGCCAGG CCCCAGGCAA GGGGCTGGAC TGGGTGGCAG TTATATCATA TGATGGAGCC      120
AATAAATACG ACGCAGACTC CGTGAAGGGC CGATTACCA TCTCCAGAGA CAATTCCAAG      180
AACACGCTGT ATCTGCAAAT GAACAGCCTG GGATCTGAGG ACACGGCTGT GTATTACTGT      240
GCGAGCGAGT CCTCCGGCTC TTAATTTCGAT CTCTGGGGCC GTGGCACCCCT GGTC          294

```

(2) INFORMATION FOR SEQ ID NO:390:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:390:

```

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAGCTATGC TATGCACTGG      60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAC TGGGTGGCAA TTATGTCATA TGATGGAACC      120
AATAAATTCG ACGCAGACTC CGTGAAGGGC CGATTACCA TCTCCAGAGA CAATTCCAAG      180
AACACGCTGT ATCTGCAAAT GAACAGCCTG AGATCTGAGG ACACGGCTGT GTATTACTGT      240
GCGAGAGAGT CTTCCGGCTG GTACTTCGAT CTCTGGGGCC GTGGCACCCCT GGTC          294

```

(2) INFORMATION FOR SEQ ID NO:391:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:391:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA ATAACGTAC TCTACACTGG	60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAC TGGGTGTCAG TTATATCATA TGATGGAGCC	120
AATAAATACG ACGCAGACTC CGTGAAGGGC CGATTCACCA TCTCCAGAGA CAATTCCAAG	180
AACACGCTGT ATCTGCAAAT GAACAGCCTG AGATCTGAGG ACACGGCTGT GTATTACTGT	240
GCGAGAGAGT CCTCCGGCTG GTACTTCGAT CTCTGGGGCC GTGGCACCTT GGTC	294

(2) INFORMATION FOR SEQ ID NO:392:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:392:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAGCTATGC TATGCACTGG	60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAC TGGGTGGCAG TTTTTCATA TGATGGAACC	120
AATAAATACG ACGCAGACTC CGTGAAGGGC CGATTCACCA TCTCCAGAGA CAATTCCAAG	180
AACACGCTGT ATCTGCAAAT GAACAGCCTG AGATCTGAGG ACACGGCTGT TTATTACTGT	240
GCGAGAGAGT CTTCCGGCTG GTACTTCGAT TTCTGGGGCC GTGGCACCTT GGTC	294

(2) INFORMATION FOR SEQ ID NO:393:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:393:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAACCTATAC TATGCACTGG	60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAC TGGGTGGCAG TTTTTCATA TGATGGAACC	120
AATAAATACG ACGCAGACTC CGTGAAGGGC CGATTACCA TCTCCAGAGA CAATTCCAAG	180
AACACGCTGT ATCTGCAAAT GAACAGCCTG AGATCTGAGG ACACGGCTGT TTATTACTGT	240
GCGAGAGAGT CTTCCGGCTG GTACTTCGAT TTCTGGGGCC GTGGCACCT GGTC	294

(2) INFORMATION FOR SEQ ID NO:394:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:394:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA ATAACCTGTAC TATACACTGG	60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAC TGGGTGGCAG TTATATCATA TGATGGAGCC	120
AATAAATACG ACGCAGACTC CGTGAAGGGC CGATTACCA TCTCCAGAGA CAAATCCAAG	180
AACACGCTGT ATCTGCAAAT GAGCAGCCTG AGATCTGAAG ACACGGCTGT ATATTACTGT	240
GTGAGAGAGT CCTCCGGCTG GTACTTCGAT CTCTGGGGCC GTGGCACCT GGTC	294

(2) INFORMATION FOR SEQ ID NO:395:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 294 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:395:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAGCTATGC TATGCACTGG	60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAG TGGGTGGCAG TTATATCATA TGATGGAAGC	120
AATAAATACT ACGCAGACTC CGTGAAGGGC CGATTACCA TCTCCAGAGA CAATTCCAAG	180
AACACGCTGT ATCTGCAAAT GAACAGCCTG AGAGCTGAGG ACACGGCTGT GTATTACTGT	240
GCGAGAGGAT GGTTCGGGGA GTTATGGGAC TACTGGGGCC AGGGAACCT GGTC	294

(2) INFORMATION FOR SEQ ID NO:396:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 297 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:396:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GCAGCTATGC TATGCACTGG	60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAG TGGGTGGCAG TTATATCATA TGATGGAAGC	120
AATAAATACT ACGCAGACTC CGTGAAGGGC CGATTCACCA TCTCCAGAGA CAATTCCAAG	180
AACACGCTGT ATCTGCAAAT GAACAGCCTG GGAGCTGAGG ACACGGCTGT GTATTACTGT	240
GCGAGAGAGA GTCTCTTAAC TGGGGACTTT GACTACTGGG GCCAGGGAAC CCTGGTC	297

(2) INFORMATION FOR SEQ ID NO:397:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 312 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:397:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAGCTATGC TATGCACTGG	60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAG TGGGTGGCAG TTATATCATA TGATGGAAGC	120
AATAAATACA ACGCAGACTC CGTGAAGGGT CGATTCACCA TCTCCAGAGA CAACTCCAAG	180
AACACGCTGT ATCTGCAAAT GAACAGCCTG AGAGCTGAGG ACACGGCTGT GTATTACTGT	240
GCGAGAGGGT ACGATATTTT GACTGGTTAT TATGACCCGC TCTTTGACAA CTGGGGCCAG	300
GGAACCCTGG TC	312

(2) INFORMATION FOR SEQ ID NO:398:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 312 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:398:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAGCTATAC TATGCACTGG	60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAG TGGGTGGCAG TTATATCATA TGATGGAAGC	120
AATAAATACT ACGCAGACTC CGTGAAGGGC CGATTCACCA TCTCCAGAGA CAATTCCAAG	180
AACACGCTGT ATCTGCAAAT GAACAGCCTG AGAGCTGAGG ACACGGCTGT GTATTACTGT	240
GCGAGAGGGT ACGATATTTT GACTGGTTAT TATGACCCGC TCTTTGACAA CTGGGGCCAG	300
GGAACCCTGG TC	312

(2) INFORMATION FOR SEQ ID NO:399:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:399:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAGCTATGC TATGCACTGG	60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAG TGGGTGGCAG TTATATCATA TGATGGAAGC	120
AATAAATACT ACGCAGACTC CGTGAAGGGC CGATTCACCA TCTCCAGAGA CAGTTCCAAG	180
AACACGCTGT ATCTGCAAAT GAACAGCCTG AGAGCTGAGG ACACGGCTGT GTATTACTGT	240
GCGAGAGATC AGGCGGCGTA TAGTGGCTAC GGGTCGGGGG GTATGGACGT CTGGGGCCAA	300
GGGACCACGG TC	312

(2) INFORMATION FOR SEQ ID NO:400:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:400:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAGCTATAC TATGCACTGG	60
GTCCGCCAGG CTCCAGGCAA GGGGCTGGAG TGGGTGGCAG TTATATCATA TGATGGAAGC	120
AATAAATACT ACGCAGACTC CGTGAAGGGC CGATTCACCA TCTCCAGAGA CAATTCCAAG	180

AACACGCTGT ATCTGCAAAT GAACAGCCTG AGAGCTGAGG ACACGGCTGT GTATTACTGT 240
 GCGAGCCATT ACTATGGTTC GGGGAGTTAT AGCTACTACG GTATGGACGT CTGGGGCCAA 300
 GGGACCACGG TC 312

(2) INFORMATION FOR SEQ ID NO:401:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 312 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:401:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAGCTATAC TATGCACTGG 60
 GTCCGCCAGG CTCCAGGCAA GGGGCTGGGG TGGGTGGCAG TTATATCATA TGATGGAAGC 120
 AATAAATACT ACGCAGACTC CGTGAAGGGC CGATTCACCA TCTCCAGAGA CAATTCCAAG 180
 AACACGCTGT ATCTGCAAAT GAACAGCCTG AGAGCTGAGG ACACGGCTGT GTATTACTGT 240
 GCGAGCCATT ACTATGGTTC GGGGAGTTAT AGCTACTACG GTATGGACGT CTGGGGCCAA 300
 GGGACCACGG TC 312

(2) INFORMATION FOR SEQ ID NO:402:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 306 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:402:

TCCCTGAGAC TCTCCTGTGC AGCCTCTGGA TTCACCTTCA GTAGCTATGC TATGCACTGG 60
 GTCCGCCAGG CTCCAGGCAA GGGGCTGGAG TGGGTGGCAG TTATATTATA TGACGAAAGC 120
 AATAAATATT ACGCAGACTC CGTGAAGGGC CGAATCACCA TCTCCAGAGA CAATTCCAAG 180
 AACACGCTGT ATCTGCAAAT GAACAGCCTG AGAGCTGAGG ACACGGCTGT GTATTACTGT 240
 GCGAGAGAGG GGA CTACTACTAC TACGGTATGG ACGTCTGGGG CCAAGGGACC 300
 ACGGTC 306

(2) INFORMATION FOR SEQ ID NO:403:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 246 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:403:

TCGGTGAAGG TCTCCTGCAA GGCTTCTGGA GGCACCTTCA GCAGCTATGC TATCAGCTGG	60
GTGCGACAGG CCCCTGGACA AGGGCTTGAG TGGATGGGAA GGATCATCCC TATCCTTGGT	120
ATAGCAAAC TACGCACAGAA GTTCCAGGGC AGAGTCACGA TTACCGCGGA CAAATCCACG	180
AGCACAGCCT ACATGGAGCT GAGCAGCCTG AGATCTGAGG ACACGGCCGT GTATTACTGT	240
GCGAGA	246

(2) INFORMATION FOR SEQ ID NO:404:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 300 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:404:

TCGGTGAAGG TCTCCTGCAA GGCTTCTGGA GGCACCTTCA GCACCTATGC TATCACCTGG	60
GTGCGACAGG CCCCTGGACA AGGGCTTGAG TGGATGGGAA AGATCATCCC TATCTTTGGT	120
ATAGCAAAC TACGCACAGAA GTTCCAGGGC AGAGTCACGA TTACCGCGGA CAAATCCACG	180
AGCACAGCCT ACATGGAGCT GACCAGCCTG AGATCTGAGG ACACGGCCGT GTATTACTGT	240
GCGAGAGACG AGACTGGGGA TCTCGGTGCT TTTGATATCT GGGGCCAAGG GACAATGGTC	300

(2) INFORMATION FOR SEQ ID NO:405:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 243 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:405:

```

ACCCTGTCCC TCACCTGCGC TGTCTATGGT GGGTCCTTCA GTGGTTACTA CTGGAGCTGG      60
ATCCGCCAGC CCCCAGGGAA GGGGCTGGAG TGGATTGGGG AAATCAATCA TAGTGGAAGC      120
ACCAACTACA ACCCGTCCCT CAAGAGTCGA GTCACCATAT CAGTAGACAC GTCCAAGAAC      180
CAGTTCTCCC TGAAGCTGAG CTCTGTGACC GCCGCGGACA CGGCTGTGTA TTACTGTGCG      240
AGA                                                                    243

```

(2) INFORMATION FOR SEQ ID NO:406:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:406:

```

ACCCTGTCCC TCACCTGCGC TGTCTATGGT GGGTCCTTCA GTGGTTACTA CTGGAGCTGG      60
ATCCGCCAGC CCCCAGGGAA GGGGCTGGAG TGGATTGGGG AAATCAATCA TAGTGGAAGC      120
ACCAACTACA ACCCGTCCCT CAAGAGTCGA GTCACCATAT CAGTAGACAC GTCCAAGAAC      180
CAGTTCTCCC TGAAGCTGAG CTCTGTGACC GCCGCGGACA CGGCTGTGTA TTACTGTGCG      240
AGGAACTTAT TTTTGGACTA CTGGGGCCAG GGAACCTGG TC                          282

```

(2) INFORMATION FOR SEQ ID NO:407:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 285 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:407:

```

ACCCTGTCCC TCACCTGCGC TGTCTATGGT GGGTCCTTCA GTGGTTACTA CTGGAGCTGG      60
ATCCGCCAGC CCCCAGGGAA GGGGCTGGAG TGGATTGGGG AAATCAATCA TAGTGGAAGC      120
ACCAACTACA ACCCGTCCCT CAAGAGTCGA GTCACCATAT CAGTAGACAC GTCCAAGAAC      180
CAGTTCTCCC TGAAGCTGAG CTCTGTGACC GCCGCGGACA CGGCTGTGTA TTACTGTGCG      240
AGGGCAGCTA ACTGGTTTGA CTACTGGGGC CAGGGAACCC TGGTC                      285

```


(2) INFORMATION FOR SEQ ID NO:408:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 297 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:408:

```

ACCCTGTCCC TCACCTGCGC TGTCTATGGT GGGTCCTTCA GTGGTTACTA CTGGAGCTGG      60
ATCCGCCAGC CCCCAGGGAA GGGGCTGGAG TGGATTGGGG AAATCAATCA TAGTGGAAGC      120
ACCAACTACA ACCCGTCCCT CAAGAGTCGA GTCACCATAT CAGTAGACAC GTCCAAGAAC      180
CAGTTCTCCC TGAAGCTGAG CTCTGTGACC GCCGCGGACA CGGCTGTGTA TTA CTGTGCA      240
GAGAGTGAGG GATGGGGATG GGACTACTTT GACTACTGGG GCCAGGGAAC CCTGGTC      297

```

(2) INFORMATION FOR SEQ ID NO:409:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3122 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:409:

```

AATTAGCGGC CGCTGTCGAC AAGCTTCGAA TTCAGTATCG ATGTGGTACC TGGATCCTCG      60
AGTGCGGCCG CAGTATGCAA AAAAAAGCCC GCTCATTAGG CGGGCTCTTG GCAGAACATA      120
TCCATCGCGT CCGCCATCTC CAGCAGCCGC ACGCGGCGCA TCTCGGGCAG CGTTGGGTCC      180
TGGCCACGGG TGCGCATGAT CGTGCTCCTG TCGTTGAGGA CCCGGCTAGG CTGGCGGGGT      240
TGCCTTACTG GTTAGCAGAA TGAATCACCG ATACGCGAGC GAACGTGAAG CGACTGCTGC      300
TGCAAAACGT CTGCGACCTG AGCAACAACA TGAATGGTCT TCGGTTTCCG TGTTTCGTAA      360
AGTCTGGAAA CGCGGAAGTC AGCGCCCTGC ACCATTATGT TCCGGATCTG CATCGCAGGA      420
TGCTGCTGGC TACCCTGTGG AACACCTACA TCTGTATTAA CGAAGCGCTG GCATTGACCC      480
TGAGTGATTT TTCTCTGGTC CCGCCGCATC CATAACGCCA GTTGTTTACC CTCACAACGT      540
TCCAGTAACC GGGCATGTTC ATCATCAGTA ACCCGTATCG TGAGCATCCT CTCTCGTTTC      600
ATCGGTATCA TTACCCCAT GAACAGAAAT TCCCCTTAC ACGGAGGCAT CAAGTGACCA      660

```

AACAGGAAAA	AACCGCCCTT	AACATGGCCC	GCTTTATCAG	AAGCCAGACA	TTAACGCTTC	720
TGGAGAAACT	CAACGAGCTG	GACGCGGATG	AACAGGCAGA	CATCTGTGAA	TCGCTTCACG	780
ACCACGCTGA	TGAGCTTTAC	CGCAGCTGCC	TCGCGCGTTT	CGGTGATGAC	GGTGAAAACC	840
TCTGACACAT	GCAGCTCCCG	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	900
GACAAGCCCG	TCAGGGCGCG	TCAGCGGGTG	TTGGCGGGTG	TCGGGGCGCA	GCCATGACCC	960
AGTCACGTAG	CGATAGCGGA	GTGTATACTG	GCTTAACTAT	GCGGCATCAG	AGCAGATTGT	1020
ACTGAGAGTG	CACCATATGC	GGTGTGAAAT	ACCGCACAGA	TGCGTAAGGA	GAAAATACCG	1080
CATCAGGCGC	TCTTCCGCTT	CCTCGCTCAC	TGACTCGCTG	CGCTCGGTCTG	TTCGGCTGCG	1140
GCGAGCGGTA	TCAGCTCACT	CAAAGGCGGT	AATACGGTTA	TCCACAGAAT	CAGGGGATAA	1200
CGCAGGAAAG	AACATGTGAG	CAAAAGGCCA	GCAAAAGGCC	AGGAACCGTA	AAAAGGCCGC	1260
GTTGCTGGCG	TTTTTCCATA	GGCTCCGCCC	CCCTGACGAG	CATCACAAAA	ATCGACGCTC	1320
AAGTCAGAGG	TGGCGAAACC	CGACAGGACT	ATAAAGATAC	CAGGCGTTTC	CCCCTGGAAG	1380
CTCCCTCGTG	CGCTCTCCTG	TTCCGACCCT	GCCGCTTACC	GGATACCTGT	CCGCCTTTCT	1440
CCCTTCGGGA	AGCGTGCGC	TTTCTCATAG	CTCACGCTGT	AGGTATCTCA	GTTCGGTGTA	1500
GGTCGTTCGC	TCCAAGCTGG	GCTGTGTGCA	CGAACCCCCC	GTTACAGCCG	ACCGCTGCGC	1560
CTTATCCGGT	AACTATCGTC	TTGAGTCCAA	CCCGGTAAGA	CACGACTTAT	CGCCACTGGC	1620
AGCAGCCAGG	CGCGCCTTGG	CCTAAGAGGC	CACTGGTAAC	AGGATTAGCA	GAGCGAGGTA	1680
TGTAGGCGGT	GCTACAGAGT	TCTTGAAGTG	GTGGCCTAAC	TACGGCTACA	CTAGAAGGAC	1740
AGTATTTGGT	ATCTGCGCTC	TGCTGAAGCC	AGTTACCTTC	GGAAAAAGAG	TTGGTAGCTC	1800
TTGATCCGGC	AAACAAACCA	CCGCTGGTAG	CGGTGGTTTT	TTTGTTTGCA	AGCAGCAGAT	1860
TACGCGCAGA	AAAAAAGGAT	CTCAAGAAGA	TCCTTTGATC	TTTTCTACGG	GGTCTGACGC	1920
TCAGTGGAAC	GAAAACTCAC	GTTAAGGGAT	TTTGGTCATG	AGATTATCAA	AAAGGATCTT	1980
CACCTAGATC	CTTTTAAATT	AAAAATGAAG	TTTTAAATCA	ATCTAAAGTA	TATATGAGTA	2040
AACTTGGTCT	GACAGTTACC	AATGCTTAAT	CAGTGAGGCA	CCTATCTCAG	CGATCTGTCT	2100
ATTCGTTCA	TCCATAGTTG	CCTGACTCCC	CGTCGTGTAG	ATAACTACGA	TACGGGAGGG	2160
CTTACCATCT	GGCCCCAGTG	CTGCAATGAT	ACCGCGAGAC	CCACGCTCAC	CGGCTCCAGA	2220
TTTATCAGCA	ATAAACCAGC	CAGCCGGAAG	GGCCGAGCGC	AGAAGTGGTC	CTGCAACTTT	2280
ATCCGCCTCC	ATCCAGTCTA	TTAATTGTTG	CCGGGAAGCT	AGAGTAAGTA	GTTCCGCCAGT	2340
TAATAGTTTG	CGCAACGTTG	TTGCCATTGC	TGCAGGCATC	GTGGTGTAC	GCTCGTCGTT	2400
TGGTATGGCT	TCATTACAGCT	CCGGTTCCCA	ACGATCAAGG	CGAGTTACAT	GATCCCCCAT	2460

GTTGTGCAAA AAAGCGGTTA GCTCCTTCGG TCCTCCGATC GTTGTGAGAA GTAAGTTGGC 2520
CGCAGTGTTA TCACTCATGG TTATGGCAGC ACTGCATAAT TCTCTTACTG TCATGCCATC 2580
CGTAAGATGC TTTTCTGTGA CTGGTGAGTA CTCAACCAAG TCATTCTGAG AATAGTGTAT 2640
GCGGCGACCG AGTTGCTCTT GCCCGGCGTC AACACGGGAT AATACCGCGC CACATAGCAG 2700
AACTTTAAAA GTGCTCATCA TTGGAAAACG TTCTTCGGGG CGAAAACCTCT CAAGGATCTT 2760
ACCGCTGTTG AGATCCAGTT CGATGTAACC CACTCGTGCA CCCAACTGAT CTTCAGCATC 2820
TTTTACTTTC ACCAGCGTTT CTGGGTGAGC AAAACAGGA AGGCAAAATG CCGCAAAAAA 2880
GGGAATAAGG GCGACACGGA AATGTTGAAT ACTCATACTC TTCCTTTTTT AATATTATTG 2940
AAGCATTTAT CAGGGTTATT GTCTCATGAG CGGATACATA TTTGAATGTA TTTAGAAAAA 3000
TAAACAAATA GGGGTTCCGC GCACATTTCC CCGAAAAGTG CCACCTGACG TCTAAGAAAC 3060
CATTATTATC ATGACATTAA CCTATAAAAA TAGGCGTATC ACGAGGCCCT TTCGTCTTCA 3120
AG 3122
